

Blind Fingerprinting

Ying Wang and Pierre Moulin

Abstract

We study blind fingerprinting, where the host sequence into which fingerprints are embedded is partially or completely unknown to the decoder. This problem relates to a multiuser version of the Gel'fand-Pinsker problem. The number of colluders and the collusion channel are unknown, and the colluders and the fingerprint embedder are subject to distortion constraints.

We propose a conditionally constant-composition random binning scheme and a universal decoding rule and derive the corresponding false-positive and false-negative error exponents. The encoder is a stacked binning scheme and makes use of an auxiliary random sequence. The decoder is a *maximum doubly-penalized mutual information decoder*, where the significance of each candidate coalition is assessed relative to a threshold that trades off false-positive and false-negative error exponents. The penalty is proportional to coalition size and is a function of the conditional type of host sequence. Positive exponents are obtained at all rates below a certain value, which is therefore a lower bound on public fingerprinting capacity. We conjecture that this value is the public fingerprinting capacity. A simpler threshold decoder is also given, which has similar universality properties but also lower achievable rates. An upper bound on public fingerprinting capacity is also derived.

Index Terms. Fingerprinting, traitor tracing, watermarking, data hiding, randomized codes, universal codes, method of types, maximum mutual information decoder, minimum equivocation decoder, channel coding with side information, random binning, capacity, error exponents, multiple access channels, model order selection.

Y. Wang is with Qualcomm, Bedminster, NJ. P. Moulin is with the ECE Department, the Coordinated Science Laboratory, and the Beckman Institute at the University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. Email: moulin@ifp.uiuc.edu. This work was supported by NSF under grants CCR 03-25924, CCF 06-35137 and CCF 07-29061. Part of this work was presented at ISIT'06 in Seattle, WA.

I. INTRODUCTION

Content fingerprinting finds applications to document protection for multimedia distribution, broadcasting, and traitor tracing [1]–[4]. A coverttext—image, video, audio, or text—is to be distributed to many users. A fingerprint, a mark unique to each user, is embedded into each copy of the coverttext. In a collusion attack, several users may combine their copies in an attempt to “remove” their fingerprints and to forge a pirated copy. The distortion between the pirated copy and the colluding copies is bounded by a certain tolerance level. To trace the forgery back to the coalition members, we need fingerprinting codes that can reliably identify the fingerprints of those members. Essentially, from a communication viewpoint, the fingerprinting problem is a multiuser version of the watermarking problem [5]–[10]. For watermarking, the attack is by one user and is based on one single copy, whereas for fingerprinting, the attack is modeled as a multiple-access channel (MAC). The coverttext plays the role of side information to the encoder and possibly to the decoder.

Depending on the availability of the original coverttext to the decoder, there are two basic versions of the problem: private and public. In the *private fingerprinting* setup, the coverttext is available to both the encoder and decoder. In the *public fingerprinting* setup, the coverttext is available to the encoder but not to the decoder, and thus decoding performance is generally worse. However public fingerprinting presents an important advantage over private fingerprinting, in that it does not require the vast storage and computational resources that are needed for media registration in a large database. For example, a DVD player could detect fingerprints from a movie disc and refuse to play it if fingerprints other than the owner’s are present. Or Web crawling programs can be used to automatically search for unauthorized content on the Internet or other public networks [3].

The scenario considered in this paper is one where a degraded version S^d of each host symbol S is available to the decoder. Private and public fingerprinting are obtained as special cases with $S^d = S$ and $S^d = \emptyset$, respectively. We refer to this scenario as either *blind* or *semiprivate fingerprinting*. The motivation is analogous to semiprivate watermarking [11], where some information about the host signal is provided to the receiver in order to improve decoding performance. This may be necessary to guarantee an acceptable performance level when the number of colluders is large.

The capacity and reliability limits of *private fingerprinting* have been studied in [7]–[10]. The decoder of [10] is a variation of Liu and Hughes’ minimum equivocation decoder [12], accounting for the presence of side information and for the fact that the number of channel inputs is unknown. Two basic types of decoders are of interest: detect-all and detect-one. The *detect-all* decoder aims to catch all members of

the coalition and an error occurs if some colluder escapes detection. The *detect-one* decoder is content with catching at least one of the culprits and an error occurs only when none of the colluders is identified. A third type of error (arguably the most damaging one) is a *false positive*, by which the decoder accuses an innocent user.

In the same way as fingerprinting is related to the MAC problem, blind fingerprinting is related to a multiuser extension of the Gel'fand-Pinsker problem. The capacity region for the latter problem is unknown. An inner region, achievable using random binning, was given in [13].

This paper derives random-coding exponents and an upper bound on detect-all capacity for semiprivate fingerprinting. Neither the encoder nor the decoder know the number of colluders. The collusion channel has arbitrary memory but is subject to a distortion constraint between the pirated copy and the colluding copies. Our fingerprinting scheme uses random binning because, unlike in the private setup, the availability of side information to the encoder and decoder is asymmetric. To optimize the error exponents, we propose an extension of the *stacked-binning* scheme that was developed for single-user channel coding with side information [11]. Here the codebook consists of a stack of variable-size codeword-arrays indexed by the conditional type of the covertext sequence. The decoder is a *minimum doubly-penalized equivocation* (M2PE) decoder or equivalently, a *maximum doubly-penalized mutual information* (M2PMI) decoder.

The proposed fingerprinting system is universal in that it can cope with unknown collusion channels and unknown number of colluders, as in the private fingerprinting setup of [10]. A tunable parameter Δ trades off false-positive and false-negative error exponents. The derivation of these exponents combines techniques from [10] and [11]. A preliminary version of our work, assuming a fixed number of colluders, was given in [14], [15].

A. Organization of This Paper

A mathematical statement of our generic fingerprinting problem is given in Sec. II, together with the basic definitions of error probabilities, capacity, error exponents, and fair coalitions. Sec. III presents our random coding scheme. Sec. IV presents a simple but suboptimal decoder that compares empirical mutual information scores between received data and individual fingerprints, and outputs a guilty decision whenever the score exceeds a certain tunable threshold. Sec. V presents a joint decoder that assigns a penalized empirical mutual information score to candidate coalitions and selects the coalition with the highest score. Sec. VI establishes an upper bound on blind fingerprinting capacity under the detect-all criterion. Finally, conclusions are given in Sec. VII. The proofs of the theorems are given in appendices.

B. Notation

We use uppercase letters for random variables, lowercase letters for their individual values, calligraphic letters for finite alphabets, and boldface letters for sequences. We denote by \mathcal{M}^* the set of sequences of arbitrary length (including 0) whose elements are in \mathcal{M} . The probability mass function (p.m.f.) of a random variable $X \in \mathcal{X}$ is denoted by $p_X = \{p_X(x), x \in \mathcal{X}\}$. The entropy of a random variable X is denoted by $H(X)$, and the mutual information between two random variables X and Y is denoted by $I(X; Y) = H(X) - H(X|Y)$. Should the dependency on the underlying p.m.f.s be explicit, we write the p.m.f.s as subscripts, e.g., $H_{p_X}(X)$ and $I_{p_X, p_{Y|X}}(X; Y)$. The Kullback-Leibler divergence between two p.m.f.s p and q is denoted by $D(p||q)$; the conditional Kullback-Leibler divergence of $p_{Y|X}$ and $q_{Y|X}$ given p_X is denoted by $D(p_{Y|X}||q_{Y|X}|p_X) = D(p_{Y|X} p_X||q_{Y|X} p_X)$. All logarithms are in base 2 unless specified otherwise.

Denote by $p_{\mathbf{x}}$ the type, or empirical p.m.f. induced by a sequence $\mathbf{x} \in \mathcal{X}^N$. The type class $T_{\mathbf{x}}$ is the set of all sequences of type $p_{\mathbf{x}}$. Likewise, we denote by $p_{\mathbf{xy}}$ the joint type of a pair of sequences $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^N \times \mathcal{Y}^N$ and by $T_{\mathbf{xy}}$ the type class associated with $p_{\mathbf{xy}}$. The conditional type $p_{\mathbf{y}|\mathbf{x}}$ of a pair of sequences (\mathbf{x}, \mathbf{y}) is defined by $p_{\mathbf{xy}}(x, y)/p_{\mathbf{x}}(x)$ for all $x \in \mathcal{X}$ such that $p_{\mathbf{x}}(x) > 0$. The conditional type class $T_{\mathbf{y}|\mathbf{x}}$ given \mathbf{x} , is the set of all sequences $\tilde{\mathbf{y}}$ such that $(\mathbf{x}, \tilde{\mathbf{y}}) \in T_{\mathbf{xy}}$. We denote by $H(\mathbf{x})$ the empirical entropy of the p.m.f. $p_{\mathbf{x}}$, by $H(\mathbf{y}|\mathbf{x})$ the empirical conditional entropy, and by $I(\mathbf{x}; \mathbf{y})$ the empirical mutual information for the joint p.m.f. $p_{\mathbf{xy}}$.

We use the calligraphic fonts \mathcal{P}_X and $\mathcal{P}_X^{[N]}$ to represent the set of all p.m.f.s and all empirical p.m.f.'s, respectively, on the alphabet \mathcal{X} . Likewise, $\mathcal{P}_{Y|X}$ and $\mathcal{P}_{Y|X}^{[N]}$ denote the set of all conditional p.m.f.s and all empirical conditional p.m.f.'s on the alphabet \mathcal{Y} . A special symbol \mathcal{W}_K will be used to denote the feasible set of collusion channels $p_{Y|X_1, \dots, X_K}$ that can be selected by a size- K coalition.

Mathematical expectation is denoted by the symbol \mathbb{E} . The shorthands $a_N \doteq b_N$ and $a_N \stackrel{\cdot}{\leq} b_N$ denote asymptotic relations in the exponential scale, respectively $\lim_{N \rightarrow \infty} \frac{1}{N} \log \frac{a_N}{b_N} = 0$ and $\limsup_{N \rightarrow \infty} \frac{1}{N} \log \frac{a_N}{b_N} \leq 0$. We define $|t|^+ \triangleq \max(t, 0)$ and $\exp_2(t) \triangleq 2^t$. The indicator function of a set \mathcal{A} is denoted by $\mathbb{1}_{\{x \in \mathcal{A}\}}$. Finally, we adopt the convention that the minimum of a function over an empty set is $+\infty$ and the maximum of a function over an empty set is 0.

II. STATEMENT OF THE PROBLEM

A. Overview

Our model for blind fingerprinting is diagrammed in Fig. 1. Let \mathcal{S} , \mathcal{X} , and \mathcal{Y} be three finite alphabets. The coartext sequence $\mathbf{S} = (S_1, \dots, S_N) \in \mathcal{S}^N$ consists of N independent and identically distributed

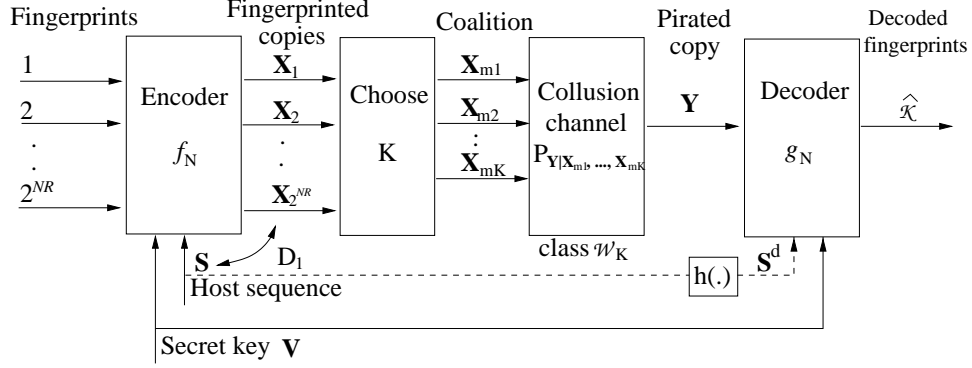


Fig. 1. Model for semiprivate (blind) fingerprinting game, where \mathbf{S}^d is a degraded version of the covertext \mathbf{S} . Private and public fingerprinting arise as special cases with $\mathbf{S}^d = \mathbf{S}$ and $\mathbf{S}^d = \emptyset$, respectively.

(i.i.d.) samples drawn from a p.m.f. $p_S(s)$, $s \in \mathcal{S}$. A random variable V taking values in an alphabet \mathcal{V}_N is shared between encoder and decoder, and not publicly revealed. The random variable V is independent of \mathbf{S} and plays the role of a cryptographic key. There are 2^{NR} users, each of which receives a fingerprinted copy:

$$\mathbf{X}_m = f_N(\mathbf{S}, V, m), \quad 1 \leq m \leq 2^{NR}, \quad (2.1)$$

where $f_N : \mathcal{S}^N \times \mathcal{V}_N \times \{1, \dots, 2^{NR}\} \rightarrow \mathcal{X}^N$ is the encoding function, and m is the index of the user. The encoder binds each fingerprinted copy \mathbf{x}_m to the covertext \mathbf{s} via a distortion constraint. Let $d : \mathcal{S} \times \mathcal{X} \rightarrow \mathbb{R}^+$ be the distortion measure and $d^N(\mathbf{s}, \mathbf{x}) = \frac{1}{N} \sum_{i=1}^N d(s_i, x_i)$ the extension of this measure to length- N sequences. The code f_N is subject to the distortion constraint

$$d^N(\mathbf{s}, \mathbf{x}_m) \leq D_1 \quad 1 \leq m \leq 2^{NR}. \quad (2.2)$$

Let $\mathcal{K} \triangleq \{m_1, m_2, \dots, m_K\}$ be a coalition of K users, called colluders. No constraints are imposed on the formation of coalitions. The colluders combine their copies $\mathbf{X}_{\mathcal{K}} \triangleq \{\mathbf{X}_m, m \in \mathcal{K}\}$ to produce a pirated copy $\mathbf{Y} \in \mathcal{Y}^N$. Without loss of generality, we assume that \mathbf{Y} is generated stochastically as the output of a collusion channel $p_{\mathbf{Y}|\mathbf{X}_{\mathcal{K}}}$. Fidelity constraints are imposed on $p_{\mathbf{Y}|\mathbf{X}_{\mathcal{K}}}$ to ensure that \mathbf{Y} is “close” to the fingerprinted copies \mathbf{X}_m , $m \in \mathcal{K}$. These constraints can take the form of distortion constraints, analogously to (2.2). They are formulated below and result in the definition of a feasible class \mathcal{W}_K of attacks.

The decoder knows neither K nor $p_{\mathbf{Y}|\mathbf{X}_{\mathcal{K}}}$ selected by the K colluders and has access to the pirated copy \mathbf{Y} , the secret key V , as well as to \mathbf{S}^d , a degraded version of the host \mathbf{S} . To simplify the exposition, the

degradation arises via a deterministic symbolwise mapping $h : \mathcal{S} \rightarrow \mathcal{S}^d$. The sequence $s^d = h(s)$ could represent a coarse version of s , or some other features of s . Two special cases are *private fingerprinting* where $\mathcal{S}^d = \mathcal{S}$, and *public fingerprinting* where $\mathcal{S}^d = \emptyset$. The decoder produces an estimate

$$\hat{\mathcal{K}} = g_N(\mathbf{Y}, \mathbf{S}^d, V) \quad (2.3)$$

of the coalition. A possible decision is the empty set, $\hat{\mathcal{K}} = \emptyset$, which is the reasonable choice when an accusation would be unreliable. To summarize, we have

Definition 2.1: A randomized rate- R length- N fingerprinting code (f_N, g_N) with embedding distortion D_1 is a pair of encoder mapping $f_N : \mathcal{S}^N \times \mathcal{V}_N \times \{1, 2, \dots, 2^{NR}\} \rightarrow \mathcal{X}^N$ and decoder mapping $g_N : \mathcal{Y}^N \times (\mathcal{S}^d)^N \times \mathcal{V}_N \rightarrow \{1, 2, \dots, 2^{NR}\}^*$.

The randomization is via the secret key V and can take the form of permutations of the symbol positions $\{1, 2, \dots, N\}$, permutations of the 2^{NR} fingerprint assignments, and an auxiliary time-sharing sequence, as in [6]—[10], [16].

We now state the attack models and define the error probabilities, capacities, and error exponents.

B. Collusion Channels

The conditional type $p_{\mathbf{Y}|\mathbf{x}_{\mathcal{K}}}$ is a random variable whose conditional distribution given $\mathbf{x}_{\mathcal{K}}$ depends on the collusion channel $p_{\mathbf{Y}|\mathbf{x}_{\mathcal{K}}}$. Our fidelity constraint on the coalition is of the general form

$$Pr[p_{\mathbf{Y}|\mathbf{x}_{\mathcal{K}}} \in \mathcal{W}_K] = 1, \quad (2.4)$$

where \mathcal{W}_K is a convex subset of $\mathcal{P}_{\mathbf{Y}|\mathbf{x}_{\mathcal{K}}}$. That is, the empirical conditional p.m.f. of the pirated copy given the marked copies is restricted. Examples of \mathcal{W}_K are given in [10], including hard distortion constraints on the coalition:

$$\mathcal{W}_K = \left\{ p_{\mathbf{Y}|\mathbf{x}_{\mathcal{K}}} : \sum_{x_{\mathcal{K}}, y} p_{X_{\mathcal{K}}}(x_{\mathcal{K}}) p_{\mathbf{Y}|\mathbf{x}_{\mathcal{K}}}(y|\mathbf{x}_{\mathcal{K}}) \mathbb{E}_{\phi} d_2(\phi(x_{\mathcal{K}}), y) \leq D_2 \right\} \quad (2.5)$$

where $\phi : \mathcal{X}^K \rightarrow \mathcal{S}$ is a (possible randomized) permutation-invariant estimator $\hat{S} = \phi(X_{\mathcal{K}})$ of each host signal sample based on the corresponding marked samples; $d_2 : \mathcal{S} \rightarrow \mathcal{Y}$ is the coalition's distortion function; $p_{X_{\mathcal{K}}}$ is a reference p.m.f.; and D_2 is the maximum allowed distortion. Another possible choice for \mathcal{W}_K is obtained using the Boneh-Shaw constraint [1], [10].

Fair Coalitions. Denote by π a permutation of the elements of \mathcal{K} . The set of fair, feasible collusion channels is the subset of \mathcal{W}_K consisting of permutation-invariant channels:

$$\mathcal{W}_K^{fair} = \{p_{\mathbf{Y}|\mathbf{x}_{\mathcal{K}}} \in \mathcal{W}_K : p_{\mathbf{Y}|\mathbf{x}_{\pi(\mathcal{K})}} = p_{\mathbf{Y}|\mathbf{x}_{\mathcal{K}}}, \forall \pi\}. \quad (2.6)$$

The collusion channel $p_{\mathbf{Y}|\mathbf{X}_K}$ is said to be fair if $Pr[p_{\mathbf{Y}|\mathbf{X}_K} \in \mathcal{W}_K^{fair}] = 1$. For any fair collusion channel, the conditional type $p_{\mathbf{Y}|\mathbf{x}_K}$ is invariant to permutations of the colluders.

Strongly exchangeable collusion channels [7]. Now denote by π a permutation of the samples of a length- N sequence. For strongly exchangeable channels, $p_{\mathbf{Y}|\mathbf{x}_K}(\pi\mathbf{y}|\pi\mathbf{x}_K)$ is independent of π , for every $(\mathbf{x}_K, \mathbf{y})$. The channel is defined by a probability assignment $Pr[T_{\mathbf{y}|\mathbf{x}_K}]$ on the conditional type classes. The distribution of \mathbf{Y} conditioned on $\mathbf{Y} \in T_{\mathbf{y}|\mathbf{x}_K}$ is uniform:

$$p_{\mathbf{Y}|\mathbf{x}_K}(\tilde{\mathbf{y}}|\mathbf{x}_K) = \frac{Pr[T_{\mathbf{y}|\mathbf{x}_K}]}{|T_{\mathbf{y}|\mathbf{x}_K}|}, \quad \forall \tilde{\mathbf{y}} \in T_{\mathbf{y}|\mathbf{x}_K}. \quad (2.7)$$

C. Error Probabilities

Let \mathcal{K} be the actual coalition and $\hat{\mathcal{K}} = g_N(\mathbf{Y}, \mathbf{S}^d, V)$ the decoder's output. The three error probabilities of interest in this paper are the probability of false positives (one or more innocent users are accused),

$$P_{FP}(f_N, g_N, p_{\mathbf{Y}|\mathbf{x}_K}) = Pr[\hat{\mathcal{K}} \setminus \mathcal{K} \neq \emptyset],$$

the probability of failing to catch a single colluder,

$$P_e^{one}(f_N, g_N, p_{\mathbf{Y}|\mathbf{x}_K}) = Pr[\hat{\mathcal{K}} \cap \mathcal{K} = \emptyset],$$

and the probability of failing to catch the full coalition:

$$P_e^{all}(f_N, g_N, p_{\mathbf{Y}|\mathbf{x}_K}) = Pr[\mathcal{K} \not\subseteq \hat{\mathcal{K}}].$$

These three probabilities are obtained by averaging over \mathbf{S} , V , and the output of the collusion channel $p_{\mathbf{Y}|\mathbf{x}_K}$. In each case the worst-case probability is denoted by

$$P_e(f_N, g_N, \mathcal{W}_K) = \max_{p_{\mathbf{Y}|\mathbf{x}_K}} P_e(f_N, g_N, p_{\mathbf{Y}|\mathbf{x}_K}) \quad (2.8)$$

where P_e denotes either P_{FP} , P_e^{one} or P_e^{all} , and the maximum is over all feasible collusion channels, i.e., such that (2.4) holds.

D. Capacity and Random-Coding Exponents

Definition 2.2: A rate R is achievable for embedding distortion D_1 , collusion class \mathcal{W}_K , and **detect-one** criterion if there exists a sequence of $(N, \lceil 2^{NR} \rceil)$ randomized codes (f_N, g_N) with maximum embedding distortion D_1 , such that both $P_{e,N}^{one}(f_N, g_N, \mathcal{W}_K)$ and $P_{FP,N}(f_N, g_N, \mathcal{W}_K)$ vanish as $N \rightarrow \infty$.

Definition 2.3: A rate R is achievable for embedding distortion D_1 , collusion class \mathcal{W}_K , and **detect-all** criterion if there exists a sequence of $(N, \lceil 2^{NR} \rceil)$ randomized codes (f_N, g_N) with maximum embedding distortion D_1 , such that both $P_{e,N}^{all}(f_N, g_N, \mathcal{W}_K)$ and $P_{FP,N}(f_N, g_N, \mathcal{W}_K)$ vanish as $N \rightarrow \infty$.

Definition 2.4: Fingerprinting capacities $C^{one}(D_1, \mathcal{W}_K)$ and $C^{all}(D_1, \mathcal{W}_K)$ are the suprema of all achievable rates with respect to the detect-one and detect-all criteria, respectively.

For random codes the error exponents corresponding to (2.8) are defined as

$$E^{\{one, all, FP\}}(R, D_1, \mathcal{W}_K) = \liminf_{N \rightarrow \infty} \left[-\frac{1}{N} \log P_e^{\{one, all, FP\}}(f_N, g_N, \mathcal{W}_K) \right]. \quad (2.9)$$

We have $C^{all}(D_1, \mathcal{W}_K) \leq C^{one}(D_1, \mathcal{W}_K)$ and $E^{all}(R, D_1, \mathcal{W}_K) \leq E^{one}(R, D_1, \mathcal{W}_K)$ because an error event for the detect-one problem is also an error event for the detect-all problem.

III. OVERVIEW OF RANDOM-CODING SCHEME

A brief overview of our scheme is given in this section. The decoders will be specified later. The scheme is designed to achieve a false-positive error exponent equal to Δ and assumes a nominal value K_{nom} for coalition size. Two arbitrarily large integers L_w and L_u are selected, defining alphabets $\mathcal{W} = \{1, 2, \dots, L_w\}$ and $\mathcal{U} = \{1, 2, \dots, L_u\}$, respectively. The parameters $\Delta, K_{nom}, L_w, L_u$ are used to identify a certain optimal type class $T_{\mathbf{w}}^*$ and conditional type classes $T_{U|S^d W}^*(\mathbf{s}^d, \mathbf{w})$, $T_{U|SW}^*(\mathbf{s}, \mathbf{w})$ and $T_{X|USW}^*(\mathbf{u}, \mathbf{s}, \mathbf{w})$ for every possible $(\mathbf{u}, \mathbf{s}, \mathbf{w})$. Optimality is defined relative to either the thresholding decoder of Sec. IV or the joint decoder of Sec. V. The secret key V consists of a random sequence $\mathbf{W} \in T_{\mathbf{w}^*}^*$ and the collection (3.1) of random codebooks indexed by $\mathbf{s}^d, \mathbf{w}, \lambda$.

A. Codebook

A random constant-composition code

$$\mathcal{C}(\mathbf{s}^d, \mathbf{w}, \lambda) = \{\mathbf{u}(l, m, \lambda), \quad 1 \leq l \leq 2^{N\rho(\lambda)}, \quad 1 \leq m \leq 2^{NR}\} \quad (3.1)$$

is generated for each pair of sequences $(\mathbf{s}^d, \mathbf{w}) \in (\mathcal{S}^d)^N \times T_{\mathbf{w}}^*$ and conditional type $\lambda \in \mathcal{P}_{S|S^d W}^{[N]}$ by drawing $2^{N[R+\rho(\lambda)]}$ random sequences independently and uniformly from an optimized conditional type class $T_{U|S^d W}^*(\mathbf{s}^d, \mathbf{w})$, and arranging them into an array with 2^{NR} columns and $2^{N\rho(\lambda)}$ rows. Similarly to [11] (see Fig. 2 therein), we refer to $\rho(\lambda)$ as the depth parameter of the array.

B. Encoding Scheme

Prior to encoding, a sequence $\mathbf{W} \in \mathcal{W}^N$ is drawn independently of \mathbf{S} and uniformly from $T_{\mathbf{w}}^*$, and shared with the receiver. Given (\mathbf{S}, \mathbf{W}) , the encoder determines the conditional type $\lambda = p_{\mathbf{S}|\mathbf{s}^d \mathbf{w}}$ and performs the following two steps for each user $1 \leq m \leq 2^{NR}$.

- 1) Find l such that $\mathbf{u}(l, m, \lambda) \in \mathcal{C}(\mathbf{s}^d, \mathbf{w}, \lambda) \cap T_{U|SW}^*(\mathbf{s}, \mathbf{w})$. If more than one such l exists, pick one of them randomly (with uniform distribution). Let $\mathbf{u} = \mathbf{u}(l, m, \lambda)$. If no such l can be found, generate \mathbf{u} uniformly from the conditional type class $T_{U|SW}^*(\mathbf{s}, \mathbf{w})$.
- 2) Generate \mathbf{X}_m uniformly distributed over the conditional type class $T_{X|USW}^*(\mathbf{u}, \mathbf{s}, \mathbf{w})$, and assign this marked sequence to user m .

C. Worst Collusion Channel

The fingerprinting codes used in this paper are randomly-modulated (RM) codes [10, Def. 2.2]. For such codes we have the following proposition, which is a straightforward variation of [10, Prop. 2.1] with \mathbf{S}^d in place of \mathbf{S} at the decoder.

Proposition 3.1: For any RM code (f_N, g_N) , the maximum of the error probability criteria (2.8) over all feasible $p_{\mathbf{Y}|\mathbf{X}_K}$ is achieved by a strongly exchangeable collusion channel, as defined in (2.7).

To derive error exponents for such channels, it suffices to use the following upper bound:

$$p_{\mathbf{Y}|\mathbf{X}_K}(\tilde{\mathbf{y}}|\mathbf{x}_K) = \frac{Pr[T_{\mathbf{y}|\mathbf{x}_K}]}{|T_{\mathbf{y}|\mathbf{x}_K}|} \leq \frac{1}{|T_{\mathbf{y}|\mathbf{x}_K}|} \mathbb{1}_{\{p_{\mathbf{y}|\mathbf{x}_K} \in \mathcal{H}_K\}}, \quad \forall \tilde{\mathbf{y}} \in T_{\mathbf{y}|\mathbf{x}_K} \quad (3.2)$$

which holds uniformly over all feasible probability assignments to conditional type classes $T_{\mathbf{y}|\mathbf{x}_K}$.

D. Encoding and Decoding Errors

The array depth parameter $\rho(\lambda)$ takes the form

$$\rho(\lambda) = I(\mathbf{u}; \mathbf{s}|\mathbf{s}^d, \mathbf{w}) + \epsilon$$

where \mathbf{u} is any element of $T_{U|SW}^*(\mathbf{s}, \mathbf{w})$, and $\epsilon > 0$ is an arbitrarily small number. The analysis shows that given any (\mathbf{s}, \mathbf{w}) , the probability of encoding errors vanishes doubly exponentially.

The analysis also shows that the decoding error probability is dominated by a single joint type class $T_{\mathbf{y}usw}$. Denote by $(\mathbf{y}, \mathbf{u}, \mathbf{s}, \mathbf{w})$ an arbitrary representative of that class. The normalized logarithm of the size of the array is given by

$$R + \rho(\lambda) = I(\mathbf{u}; \mathbf{y}|\mathbf{s}^d, \mathbf{w}) - \Delta,$$

and the probability of false positives vanishes as $2^{-N\Delta}$.

IV. THRESHOLD DECODER

A. Decoding

The decoder has access to $(\mathbf{y}, \mathbf{s}^d, \mathbf{w})$ but does not know the conditional type $\lambda = p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}$ realized at the encoder. The decoder evaluates the users one at a time and makes an innocent/guilty decision on each user *independently of the other users*. Specifically, the receiver outputs an estimated coalition $\hat{\mathcal{K}}$ if and only if $\hat{\mathcal{K}}$ satisfies the following condition:

$$\forall m \in \hat{\mathcal{K}} : \max_{\lambda \in \mathcal{P}_{S|S^dW}^{[N]}} \max_{1 \leq l \leq 2^{N\rho(\lambda)}} I(\mathbf{u}(l, m, \lambda); \mathbf{y}|\mathbf{s}^d\mathbf{w}) - \rho(\lambda) > R + \Delta. \quad (4.1)$$

If no such $\hat{\mathcal{K}}$ is found, the receiver outputs $\hat{\mathcal{K}} = \emptyset$. This decoder outputs all user indices whose empirical mutual information score, penalized by $\rho(\lambda)$, exceeds the threshold $R + \Delta$.

Observe that the maximizing λ in (4.1) *may* depend on m . With high probability, this event implies a decoding error. Improvements can only be obtained using a more complex joint decoder, as in Sec. V.

B. Error Exponents

Define the following set of conditional p.m.f.'s for $(XU)_{\mathcal{K}} \triangleq (X_{\mathcal{K}}, U_{\mathcal{K}})$ given (S, W) :

$$\mathcal{M}(p_{XU|SW}) = \{p_{(XU)_{\mathcal{K}}|SW} : p_{X_m U_m|SW} = p_{XU|SW}, m \in \mathcal{K}\},$$

i.e., the conditional marginal p.m.f. $p_{XU|SW}$ is the same for each $(X_m, U_m), \forall m \in \mathcal{K}$. Also define the sets

$$\begin{aligned} \mathcal{P}_{XU|SW}(p_{SW}, L_w, L_u, D_1) &= \{p_{XU|SW} : \mathbb{E}[d(S, X)] \leq D_1\}, \\ \mathcal{P}_{(XU)_{\mathcal{K}}W|S}(p_S, L_w, L_u, D_1) &= \left\{ p_{(XU)_{\mathcal{K}}W|S} = p_W \prod_{k \in \mathcal{K}} p_{X_k U_k|SW} \right. \\ &\quad \left. : p_{X_1 U_1|SW} = \dots = p_{X_K U_K|SW}, \text{ and } \mathbb{E}[d(S, X_1)] \leq D_1 \right\} \end{aligned} \quad (4.2)$$

where in (4.2) the random variables $(X_k, U_k), k \in \mathcal{K}$, are conditionally i.i.d. given (S, W) .

Define for each $m \in \mathcal{K}$ the set of conditional p.m.f.'s

$$\begin{aligned} &\mathcal{P}_{Y(XU)_{\mathcal{K}}|SW}(p_W, \tilde{p}_{S|W}, p_{XU|SW}, \mathcal{W}_K, R, L_w, L_u, m) \\ &\triangleq \left\{ \tilde{p}_{Y(XU)_{\mathcal{K}}|SW} : \tilde{p}_{(XU)_{\mathcal{K}}|SW} \in \mathcal{M}(p_{XU|SW}), \tilde{p}_{Y|X_{\mathcal{K}}} \in \mathcal{W}_K, \right. \\ &\quad \left. I_{p_W \tilde{p}_{S|W} \tilde{p}_{Y(XU)_{\mathcal{K}}|SW}}(U_m; Y|S^d W) - I_{p_W \tilde{p}_{S|W} p_{XU|SW}}(U; S|S^d W) \leq R \right\} \end{aligned} \quad (4.3)$$

and the *pseudo sphere packing exponent*

$$\begin{aligned} \tilde{E}_{psp,m}(R, p_W, \tilde{p}_{S|W}, p_{XU|SW}, \mathcal{W}_K) &= \min_{\tilde{p}_{Y(XU)_K|SW} \in \mathcal{P}_{Y(XU)_K|SW}(p_W, \tilde{p}_{S|W}, p_{XU|SW}, \mathcal{W}_K, R, L_w, L_u, m)} \\ &\quad D(\tilde{p}_{Y(XU)_K|SW} \tilde{p}_{S|W} \| \tilde{p}_{Y|X_K} p_{XU|SW}^{\mathcal{K}} p_{S|p_W}). \end{aligned} \quad (4.4)$$

Taking the maximum and minimum of $\tilde{E}_{psp,m}$ above over $m \in \mathcal{K}$, we respectively define

$$\overline{\tilde{E}}_{psp}(R, L_w, L_u, p_W, \tilde{p}_{S|W}, p_{XU|SW}, \mathcal{W}_K) = \max_{m \in \mathcal{K}} \tilde{E}_{psp,m}(R, L_w, L_u, p_W, \tilde{p}_{S|W}, p_{XU|SW}, \mathcal{W}_K) \quad (4.5)$$

$$\underline{\tilde{E}}_{psp}(R, L_w, L_u, p_W, \tilde{p}_{S|W}, p_{XU|SW}, \mathcal{W}_K) = \min_{m \in \mathcal{K}} \tilde{E}_{psp,m}(R, L_w, L_u, p_W, \tilde{p}_{S|W}, p_{XU|SW}, \mathcal{W}_K). \quad (4.6)$$

For a fair coalition ($\mathcal{W}_K = \mathcal{W}_K^{fair}$), $\tilde{E}_{psp,m}$ is independent of $m \in \mathcal{K}$, and the two expressions above coincide. Define

$$\begin{aligned} E_{psp}(R, L_w, L_u, D_1, \mathcal{W}_K) &= \max_{p_W \in \mathcal{P}_W} \min_{\tilde{p}_{S|W} \in \mathcal{P}_{S|W}} \max_{p_{XU|SW} \in \mathcal{P}_{XU|SW}(p_W, \tilde{p}_{S|W}, L_w, L_u, D_1)} \\ &\quad \tilde{E}_{psp,1}(R, L_w, L_u, p_W, \tilde{p}_{S|W}, p_{XU|SW}, \mathcal{W}_K^{fair}). \end{aligned} \quad (4.7)$$

Denote by p_W^* and $p_{XU|SW}^*$ the maximizers in (4.7), the latter to be viewed as a function of $\tilde{p}_{S|W}$. Both p_W^* and $p_{XU|SW}^*$ implicitly depend on R and \mathcal{W}_K^{fair} . Finally, define

$$\overline{E}_{psp}(R, L_w, L_u, D_1, \mathcal{W}_K) = \min_{\tilde{p}_{S|W} \in \mathcal{P}_{S|W}} \overline{\tilde{E}}_{psp}(R, L_w, L_u, p_W^*, \tilde{p}_{S|W}, p_{XU|SW}^*, \mathcal{W}_K) \quad (4.8)$$

$$\underline{E}_{psp}(R, L_w, L_u, D_1, \mathcal{W}_K) = \min_{\tilde{p}_{S|W} \in \mathcal{P}_{S|W}} \underline{\tilde{E}}_{psp}(R, L_w, L_u, p_W^*, \tilde{p}_{S|W}, p_{XU|SW}^*, \mathcal{W}_K). \quad (4.9)$$

The terminology *pseudo sphere-packing exponent* is used because despite its superficial similarity to a real sphere-packing exponent, (4.4) does not provide a fundamental asymptotic lower bound on error probability.

Theorem 4.1: The decision rule (4.1) yields the following error exponents.

- (i) The false-positive error exponent is

$$E_{FP}(R, D_1, \mathcal{W}_K, \Delta) = \Delta. \quad (4.10)$$

- (ii) The detect-all error exponent is

$$E^{all}(R, L_w, L_u, D_1, \mathcal{W}_K, \Delta) = \underline{E}_{psp}(R + \Delta, L_w, L_u, D_1, \mathcal{W}_K). \quad (4.11)$$

- (iii) The detect-one error exponent is

$$E^{one}(R, L_w, L_u, D_1, \mathcal{W}_K, \Delta) = \overline{E}_{psp}(R + \Delta, L_w, L_u, D_1, \mathcal{W}_K). \quad (4.12)$$

- (iv) A fair collusion strategy is optimal under the detect-one error criterion:

$$E^{one}(R, L_w, L_u, D_1, \mathcal{W}_K, \Delta) = E^{one}(R, L_w, L_u, D_1, \mathcal{W}_K^{fair}, \Delta).$$

- (v) The detect-one and detect-all error exponents are the same when the colluders employ a fair strategy: $E^{one}(R, L_w, L_u, D_1, \mathcal{W}_K^{fair}, \Delta) = E^{all}(R, L_w, L_u, D_1, \mathcal{W}_K^{fair}, \Delta)$.
- (vi) For $K = K_{nom}$, the supremum of all rates for which the detect-one error exponent of (4.12) is positive is

$$\begin{aligned}
C^{thr}(D_1, \mathcal{W}_K) &= C^{thr}(D_1, \mathcal{W}_K^{fair}) \\
&= \lim_{L_w, L_u \rightarrow \infty} \max_{p_W \in \mathcal{P}_W} \max_{p_{XU|SW} \in \mathcal{P}_{XU|SW}(p_W, p_{S, L_w, L_u, D_1})} \min_{p_{Y|XK} \in \mathcal{W}_K^{fair}} \\
&\quad [I(U; Y|S^d, W) - I(U; S|S^d, W)].
\end{aligned} \tag{4.13}$$

V. JOINT FINGERPRINT DECODER

The fundamental improvement over the simple thresholding strategy for decoding in Sec. IV resides in the use of a joint decoding rule. Specifically, the decoder maximizes a penalized empirical mutual information score over all possible coalitions of any size. The penalty depends on the conditional host sequence type $p_{S|S^d, W}$, as in Sec. IV, and is proportional to the size of the coalition, as in [10, Sec. V]. We call this blind fingerprint decoder the *maximum doubly-penalized mutual information* (M2PMI) decoder.

Mutual Information of k Random Variables. The mutual information of k random variables X_1, \dots, X_k is defined as the sum of their individual entropies minus their joint entropy [21, p. 57] or equivalently, the divergence between their joint distribution and the product of their marginals:

$$\begin{aligned}
\overset{\circ}{I}(X_1; \dots; X_k) &= H(X_1) + \dots + H(X_k) - H(X_1, \dots, X_k) \\
&= D(p_{X_1 \dots X_k} \| p_{X_1} \dots p_{X_k}).
\end{aligned} \tag{5.1}$$

The symbol $\overset{\circ}{I}$ is used to distinguish it from ordinary mutual information I between two random variables. Similarly one can define a conditional mutual information $\overset{\circ}{I}(X_1; \dots; X_k|Z) = \sum_i H(X_i|Z) - H(X_1, \dots, X_k|Z)$ conditioned on Z , and an empirical mutual information $\overset{\circ}{I}(\mathbf{x}_1; \dots; \mathbf{x}_k|\mathbf{z})$ between k sequences $\mathbf{x}_1, \dots, \mathbf{x}_k$, conditioned on \mathbf{z} , as the conditional mutual information with respect to the joint type of $\mathbf{x}_1, \dots, \mathbf{x}_k, \mathbf{z}$. Some properties of $\overset{\circ}{I}$ are given in [10, Sec. V.A].

Recall that $\mathbf{x}_{\mathcal{A}}$ denotes $\{\mathbf{x}_m, m \in \mathcal{A}\}$ and that the codewords in (3.1) take the form $\mathbf{u}(l, m, \lambda)$. In the following, we shall use the compact notation $(\mathbf{x}\mathbf{u})_{\mathcal{A}} \triangleq (\mathbf{x}_{\mathcal{A}}, \mathbf{u}_{\mathcal{A}})$, and

$$\mathbf{u}(l_{\mathcal{A}}, m_{\mathcal{A}}, \lambda) \triangleq \{\mathbf{u}(l_{m_1}, m_1, \lambda), \dots, \mathbf{u}(l_{m_{|\mathcal{A}|}}, m_{|\mathcal{A}|}, \lambda)\} \quad \text{for } \mathcal{A} = \{m_1, \dots, m_{|\mathcal{A}|}\}.$$

A. M2PMI Criterion

Given $\mathbf{y}, \mathbf{s}^d, \mathbf{w}$, the decoder seeks the coalition size k , the conditional host sequence type $\lambda \in \mathcal{P}_{S|S^d W}^{[N]}$, and the codewords $\mathbf{u}(l, m, \lambda)$ in $\mathcal{C}(\mathbf{s}^d, \mathbf{w}, \lambda)$ that maximize the M2PMI criterion below. The column indices $m \in \mathcal{K}$, corresponding to the decoded words form the decoded coalition $\hat{\mathcal{K}}$. If the maximizing k in (5.2) is zero, the receiver outputs $\hat{\mathcal{K}} = \emptyset$.

The *Maximum Doubly-Penalized Mutual Information* criterion is defined as

$$\max_{k \geq 0} M2PMI(k) \quad (5.2)$$

where

$$M2PMI(k) = \begin{cases} 0 & : \text{if } k = 0 \\ \max_{\lambda \in \mathcal{P}_{S|S^d W}^{[N]}} \max_{\mathbf{u}_{\mathcal{K}} \in \mathcal{C}^k(\mathbf{s}^d, \mathbf{w}, \lambda)} \left[\overset{\circ}{I}(\mathbf{u}_{\mathcal{K}}; \mathbf{y} | \mathbf{s}^d \mathbf{w}) - k(\rho(\lambda) + R + \Delta) \right] & : \text{if } k = 1, 2, \dots \end{cases} \quad (5.3)$$

B. Properties

The following lemma shows that 1) each subset of the estimated coalition is significant, and 2) any further extension of the coalition would fail a significance test. The proof parallels that of Lemma 5.1 in [10] and is therefore omitted.

Lemma 5.1: Let $\hat{\mathcal{K}}, \lambda, l_{\hat{\mathcal{K}}}$ achieve the maximum in (5.3) (5.2), i.e., $\mathbf{u}_{\hat{\mathcal{K}}} = \mathbf{u}(l_{\hat{\mathcal{K}}}, m_{\hat{\mathcal{K}}}, \lambda)$. Then for each subset of the estimated coalition $\hat{\mathcal{K}}$, we have

$$\forall \mathcal{A} \subseteq \hat{\mathcal{K}} : \overset{\circ}{I}(\mathbf{u}(l_{\mathcal{A}}, m_{\mathcal{A}}, \lambda); \mathbf{y}_{\mathbf{u}(l_{\hat{\mathcal{K}} \setminus \mathcal{A}}, m_{\hat{\mathcal{K}} \setminus \mathcal{A}}, \lambda)} | \mathbf{s}^d \mathbf{w}) > |\mathcal{A}| (\rho(\lambda) + R + \Delta). \quad (5.4)$$

Moreover, for every \mathcal{A} disjoint with $\hat{\mathcal{K}}$,

$$\overset{\circ}{I}(\mathbf{u}(l_{\mathcal{A}}, m_{\mathcal{A}}, \lambda); \mathbf{y}_{\mathbf{u}(l_{\hat{\mathcal{K}}}, m_{\hat{\mathcal{K}}}, \lambda)} | \mathbf{s}^d \mathbf{w}) \leq |\mathcal{A}| (\rho(\lambda) + R + \Delta). \quad (5.5)$$

C. Error Exponents

Define for each $\mathcal{A} \subseteq \mathcal{K}$ the set of conditional p.m.f.'s

$$\begin{aligned} & \mathcal{P}_{Y(XU)_{\mathcal{K}}|SW}(p_W, \tilde{p}_{S|W}, p_{XU|SW}, \mathcal{W}_K, R, L_w, L_u, \mathcal{A}) \\ & \triangleq \left\{ \tilde{p}_{Y(XU)_{\mathcal{K}}|SW} : \tilde{p}_{(XU)_{\mathcal{K}}|SW} \in \mathcal{M}(p_{XU|SW}), \right. \\ & \quad \left. \frac{1}{|\mathcal{A}|} \overset{\circ}{I}_{p_W \tilde{p}_{S|W} \tilde{p}_{Y(XU)_{\mathcal{K}}|SW}}(U_{\mathcal{A}}; Y_{U_{\mathcal{K} \setminus \mathcal{A}}} | S^d, W) \leq I_{p_W \tilde{p}_{S|W} p_{XU|SW}}(U; S | S^d, W) + R \right\} \quad (5.6) \end{aligned}$$

and the *pseudo sphere packing exponent*

$$\begin{aligned} \tilde{E}_{psp,\mathcal{A}}(R, L_w, L_u, p_W, \tilde{p}_{S|W}, p_{XU|SW}, \mathcal{W}_K) &= \min_{\substack{\tilde{p}_{Y(XU)_K|SW} \in \mathcal{P}_{Y(XU)_K|SW}(p_W, \tilde{p}_{S|W}, p_{XU|SW}, \mathcal{W}_K, R, L_w, L_u, \mathcal{A}) \\ D(\tilde{p}_{Y(XU)_K|SW} \tilde{p}_{S|W} \| \tilde{p}_{Y|X_K} \tilde{p}_{(XU)_K|SW} p_S | p_W)}} \end{aligned} \quad (5.7)$$

Taking the maximum ¹ and the minimum of $\tilde{E}_{psp,\mathcal{A}}$ above over all subsets \mathcal{A} of \mathcal{K} , we define

$$\overline{\tilde{E}}_{psp}(R, L_w, L_u, p_W, \tilde{p}_{S|W}, p_{XU|SW}, \mathcal{W}_K) = \tilde{E}_{psp,\mathcal{K}}(R, L_w, L_u, p_W, \tilde{p}_{S|W}, p_{XU|SW}, \mathcal{W}_K), \quad (5.8)$$

$$\underline{\tilde{E}}_{psp}(R, L_w, L_u, p_W, \tilde{p}_{S|W}, p_{XU|SW}, \mathcal{W}_K) = \min_{\mathcal{A} \subseteq \mathcal{K}} \tilde{E}_{psp,\mathcal{A}}(R, L_w, L_u, p_W, \tilde{p}_{S|W}, p_{XU|SW}, \mathcal{W}_K). \quad (5.9)$$

Now define

$$\begin{aligned} E_{psp}(R, L_w, L_u, D_1, \mathcal{W}_K) &= \max_{p_W \in \mathcal{P}_W} \min_{\tilde{p}_{S|W} \in \mathcal{P}_{S|W}} \max_{p_{XU|SW} \in \mathcal{P}_{XU|SW}} \tilde{E}_{psp,\mathcal{K}}(R, L_w, L_u, p_W, \tilde{p}_{S|W}, p_{XU|SW}, \mathcal{W}_K^{fair}). \end{aligned} \quad (5.10)$$

Denote by p_W^* and $p_{XU|SW}^*$ the maximizers in (5.10), where the latter is to be viewed as a function of $\tilde{p}_{S|W}$. Both p_W^* and $p_{XU|SW}^*$ implicitly depend on R and \mathcal{W}_K^{fair} . Finally, define

$$\overline{E}_{psp}(R, L_w, L_u, D_1, \mathcal{W}_K) = \min_{\tilde{p}_{S|W} \in \mathcal{P}_{S|W}} \overline{\tilde{E}}_{psp}(R, L_w, L_u, p_W^*, \tilde{p}_{S|W}, p_{XU|SW}^*, \mathcal{W}_K), \quad (5.11)$$

$$\underline{E}_{psp}(R, L_w, L_u, D_1, \mathcal{W}_K) = \min_{\tilde{p}_{S|W} \in \mathcal{P}_{S|W}} \underline{\tilde{E}}_{psp}(R, L_w, L_u, p_W^*, \tilde{p}_{S|W}, p_{XU|SW}^*, \mathcal{W}_K). \quad (5.12)$$

Theorem 5.2: The decision rule (5.2) yields the following error exponents.

(i) The false-positive error exponent is

$$E_{FP}(R, D_1, \mathcal{W}_K, \Delta) = \Delta. \quad (5.13)$$

(ii) The detect-all error exponent is

$$E^{all}(R, L_w, L_u, D_1, \mathcal{W}_K, \Delta) = \underline{E}_{psp}(R + \Delta, L_w, L_u, D_1, \mathcal{W}_K). \quad (5.14)$$

(iii) The detect-one error exponent is

$$E^{one}(R, L_w, L_u, D_1, \mathcal{W}_K, \Delta) = \overline{E}_{psp}(R + \Delta, L_w, L_u, D_1, \mathcal{W}_K). \quad (5.15)$$

(iv) $E^{one}(R, L_w, L_u, D_1, \mathcal{W}_K, \Delta) = E^{one}(R, L_w, L_u, D_1, \mathcal{W}_K^{fair}, \Delta)$.

(v) $E^{all}(R, L_w, L_u, D_1, \mathcal{W}_K^{fair}, \Delta) = E^{one}(R, L_w, L_u, D_1, \mathcal{W}_K^{fair}, \Delta)$.

¹ The property that \mathcal{K} achieves $\max_{\mathcal{A} \subseteq \mathcal{K}} \tilde{E}_{psp,\mathcal{A}}$ is established in the proof of Theorem 5.2, Part (iv).

- (vi) If $K = K_{nom}$, the supremum of all rates for which the error exponent of (5.15) and (5.14) are positive is

$$\begin{aligned} \underline{C}^{one}(D_1, \mathcal{W}_K) &= \underline{C}^{one}(D_1, \mathcal{W}_K^{fair}) \\ &= \lim_{L_w, L_u \rightarrow \infty} \max_{p_W \in \mathcal{P}_W} \max_{p_{(XU)_K|SW} \in \mathcal{P}_{(XU)_K|SW}(p_W, p_S, L_w, L_u, D_1)} \min_{p_{Y|X_K} \in \mathcal{W}_K^{fair}} \\ &\quad \left[\frac{1}{K} I(U_K; Y | S^d, W) - I(U; S | S^d, W) \right] \end{aligned} \quad (5.16)$$

under the “detect-one” criterion, and by

$$\begin{aligned} \underline{C}^{all}(D_1, \mathcal{W}_K) &= \lim_{L_w, L_u \rightarrow \infty} \max_{p_W \in \mathcal{P}_W} \max_{p_{(XU)_K|SW} \in \mathcal{P}_{(XU)_K|SW}(p_W, p_S, L_w, L_u, D_1)} \min_{p_{Y|X_K} \in \mathcal{W}_K} \\ &\quad \left[\min_{\mathcal{A} \subseteq \mathcal{K}} \frac{1}{|\mathcal{A}|} I(U_{\mathcal{A}}; Y | S^d, W, U_{\mathcal{K} \setminus \mathcal{A}}) - I(U; S | S^d, W) \right] \end{aligned} \quad (5.17)$$

under the “detect-all” criterion. If the colluders select a fair collusion channel, as is their collective interest, the minimization is restricted to \mathcal{W}_K^{fair} in (5.17), and then

$$\underline{C}^{all}(D_1, \mathcal{W}_K) = \underline{C}^{one}(D_1, \mathcal{W}_K).$$

For the special case of private fingerprinting ($S^d = S$), the term $I(U; S | S^d, W)$ in (5.16) is zero. Since $I(U_K; Y | S, W) \leq I((XU)_K; Y | S, W)$, it suffices to choose $L_u = |\mathcal{X}|$ and $U = X$ to achieve the maximum in (5.16). The resulting expression coincides with the capacity formula in [10, Theorem 3.2]. Similarly to the single-user case [11], when $U = X$ the binning scheme is degenerate.

D. Bounded Coalition Size

Assume now that K is known not exceed some maximum value K_{max} . The same random coding scheme can be used. In the evaluation of the M2PMI criterion of (5.2), the maximization is now limited to $0 \leq k \leq K_{max}$. In Lemma 5.1, property (5.4) holds, and property (5.5) now holds for every \mathcal{A} disjoint with $\hat{\mathcal{K}}$, and of size $|\mathcal{A}| \leq K_{max} - |\hat{\mathcal{K}}|$. Following the derivation of the error exponents in the appendix, we see that these exponents remain the same as those given by Theorem 5.2.

Blind watermarking. The case $K_{max} = 1$ represents blind watermark decoding with a guarantee that the false-positive exponent is at least equal to Δ . In this scenario, there is no need for a time-sharing sequence \mathbf{w} , and the decoder’s input \mathbf{y} is either an unwatermarked sequence ($K = 0$) or a watermarked sequence ($K = 1$). The M2PMI criterion of (5.3) reduces to

$$M2PMI(k) = \max_{\lambda} \max_{\mathbf{u} \in \mathcal{C}(s^d)} I(\mathbf{u}; \mathbf{y} | s^d) - (\rho(\lambda) + R + \Delta) \quad \text{for } k = 1.$$

The resulting false-positive and false-negative exponents are given by Δ and $E_{psp}(R + \Delta, 0, L_u, D_1, \mathcal{W}_K)$, respectively.

VI. UPPER BOUNDS ON PUBLIC FINGERPRINTING CAPACITY

Deriving public fingerprinting capacity is a challenge because the capacity region for the Gel'fand-Pinsker version of the MAC is still unknown, in fact an outer bound for this region has yet to be established. Even in the case of a MAC with side information *causally* available at the transmitter but not at the receiver, the expressions for the inner and outer capacity regions do not coincide [23]. Likewise, the expression derived below is an upper bound on public fingerprinting capacity under the detect-all criterion.

Recall the definition of the set $\mathcal{P}_{(XU)_{\mathcal{K}}W|S}(p_S, L_w, L_u, D_1)$ in (4.2), where W and U are random variables defined over alphabets $\mathcal{W} = \{1, 2, \dots, L_w\}$ and $\mathcal{U} = \{1, 2, \dots, L_u\}$, respectively. Here we define the larger set

$$\begin{aligned} \mathcal{P}_{(XU)_{\mathcal{K}}W|S}^{outer}(p_S, L_w, L_u, D_1) = & \left\{ p_{(XU)_{\mathcal{K}}W|S} = p_W \left(\prod_{k \in \mathcal{K}} p_{X_k|SW} \right) p_{U_{\mathcal{K}}|X_{\mathcal{K}}SW} : \right. \\ & \left. p_{X_1|SW} = \dots = p_{X_{\mathcal{K}}|SW}, \text{ and } \mathbb{E}[d(S, X_1)] \leq D_1 \right\} \end{aligned} \quad (6.1)$$

where $X_k, k \in \mathcal{K}$, are still conditionally i.i.d. given (S, W) but the random variables $U_k, k \in \mathcal{K}$, are generally conditionally dependent.

Define

$$\begin{aligned} \overline{\mathcal{C}}_{L_w, L_u}^{all}(D_1, \mathcal{W}_K) = & \max_{p_{(XU)_{\mathcal{K}}W|S} \in \mathcal{P}_{(XU)_{\mathcal{K}}W|S}^{outer}(p_S, L_w, L_u, D_1)} \min_{p_{Y|X_{\mathcal{K}}} \in \mathcal{W}_K} \\ & \min_{\mathcal{A} \subseteq \mathcal{K}} \frac{1}{|\mathcal{A}|} \left[I(U_{\mathcal{A}}; Y, S^d | U_{\mathcal{K} \setminus \mathcal{A}}) - I(U_{\mathcal{A}}; S | U_{\mathcal{K} \setminus \mathcal{A}}) \right]. \end{aligned} \quad (6.2)$$

Using the same derivation as in Lemma 2.1 of [11], it can be shown that $\overline{\mathcal{C}}_{L_w, L_u}^{all}(D_1, \mathcal{W}_K)$ is a nondecreasing function of L_w and L_u and converges to a finite limit. Moreover, the gap to the limit may be bounded by a polynomial function of L_w and L_u , see [11, Sec. 3.5] for a similar derivation.

Theorem 6.1: Public fingerprinting capacity is upper-bounded by

$$\overline{\mathcal{C}}^{all}(D_1, \mathcal{W}_K) = \lim_{L_w, L_u \rightarrow \infty} \overline{\mathcal{C}}_{L_w, L_u}^{all}(D_1, \mathcal{W}_K) \quad (6.3)$$

under the “detect-all” criterion.

Proof: see appendix.

We conjecture that the upper bound on capacity given by Theorem 6.1 is generally not tight. The insight here is that the upper bound remains valid if the class of encoding functions is enlarged to include feedback from the receiver: $X_{ki} = \tilde{f}_i(\mathbf{S}, M_k, Y^{i-1})$ for $1 \leq i \leq N$. It can indeed be verified that all the inequalities in the proof and the Markov chain properties hold. The question is now whether

feedback can increase public fingerprinting capacity. We conjecture the answer is yes, because feedback is known to increase MAC capacity [24].

We also make the stronger conjecture that the maximum over $p_{(XU)_K|SW}$ is achieved by a p.m.f. that decouples the components (X_k, U_k) , $k \in \mathcal{K}$, conditioned on (S, W) . If this is true, the set $\mathcal{P}_{(XU)_K W|S}^{outer}(p_S, L_w, L_u, D_1)$ in the formula (6.2) can be replaced with the smaller set $\mathcal{P}_{(XU)_K W|S}(p_S, L_w, L_u, D_1)$ of (4.2), and the random coding scheme of Sec. V is capacity-achieving.

VII. CONCLUSION

We have proposed a communication model and a random-coding scheme for blind fingerprinting. While a standard binning scheme for communication with asymmetric side information at the transmitter and the receiver may seem like a reasonable candidate, such a scheme would be unable to trade false-positive error exponents against false-negative error exponents. Our proposed binning scheme combines two ideas. The first is the use of a stacked binning scheme as in [11], which demonstrated the advantages (in terms of decoding error exponents) of selecting codewords from an array whose size depends on the conditional type of the host sequence. The second is the use of an auxiliary time-sharing random variable as in [10]. The blind fingerprint decoders of Secs. IV and V combine the advantages of both methods and provide positive error exponents for a range of code rates. The tradeoff between the two fundamental types of error probabilities is determined by the value of the parameter Δ .

APPENDIX I

PROOF OF THEOREM 4.1

We derive the error exponents for the thresholding rule (4.1). We have $\mathcal{W} = \{1, 2, \dots, L_w\}$ and $\mathcal{U} = \{1, 2, \dots, L_u\}$. Fix some arbitrarily small $\epsilon > 0$. Define for all $m \in \mathcal{K}$

$$\begin{aligned} \mathcal{P}_{Y(XU)_{\mathcal{K}}|SW}^{[N]}(p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_K, R, L_w, L_u, m) &= \left\{ p_{\mathbf{y}(\mathbf{xu})_{\mathcal{K}}|\mathbf{sw}} : p_{(\mathbf{xu})_{\mathcal{K}}|\mathbf{sw}} \in \mathcal{M}(p_{\mathbf{xu}|\mathbf{sw}}), \right. \\ &\quad \left. p_{\mathbf{y}|\mathbf{x}_{\mathcal{K}}} \in \mathcal{W}_K, I(\mathbf{u}_m; \mathbf{y}|\mathbf{s}^d \mathbf{w}) \leq \rho(p_{\mathbf{s}|\mathbf{s}^d \mathbf{w}}) + R \right\} \\ \check{E}_{psp,m,N}(R, L_w, L_u, p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_K) &= \min_{p_{\mathbf{y}(\mathbf{xu})_{\mathcal{K}}|\mathbf{sw}} \in \mathcal{P}_{Y(XU)_{\mathcal{K}}|SW}^{[N]}(p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_K, R, L_w, L_u, m)} \\ &\quad D(p_{\mathbf{y}(\mathbf{xu})_{\mathcal{K}}|\mathbf{sw}} \| p_{\mathbf{y}|\mathbf{x}_{\mathcal{K}}} p_{\mathbf{xu}|\mathbf{sw}}^{\mathcal{K}} | p_{\mathbf{sw}}), \end{aligned} \quad (\text{A.1})$$

$$\begin{aligned} \hat{E}_{psp,m,N}(R, L_w, L_u, p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_K) &= D(p_{\mathbf{s}|\mathbf{w}} \| p_S | p_{\mathbf{w}}) + \check{E}_{psp,m,N}(R, L_w, L_u, \\ &\quad p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_K) \\ &= \min_{p_{\mathbf{y}(\mathbf{xu})_{\mathcal{K}}|\mathbf{sw}} \in \mathcal{P}_{Y(XU)_{\mathcal{K}}|SW}^{[N]}(p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_K, R, L_w, L_u, m)} \\ &\quad D(p_{\mathbf{y}(\mathbf{xu})_{\mathcal{K}}|\mathbf{sw}} p_{\mathbf{s}|\mathbf{w}} \| p_{\mathbf{y}|\mathbf{x}_{\mathcal{K}}} p_{\mathbf{xu}|\mathbf{sw}}^{\mathcal{K}} p_S | p_{\mathbf{w}}), \end{aligned} \quad (\text{A.2})$$

$$\overline{\hat{E}}_{psp,N}(R, L_w, L_u, p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_K) = \max_{m \in \mathcal{K}} \hat{E}_{psp,m,N}(R, L_w, L_u, p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_K) \quad (\text{A.3})$$

$$\underline{\hat{E}}_{psp,N}(R, L_w, L_u, p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_K) = \min_{m \in \mathcal{K}} \hat{E}_{psp,m,N}(R, L_w, L_u, p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_K) \quad (\text{A.4})$$

where (A.2) is obtained by application of the chain rule for divergence. Also define

$$\begin{aligned} E_{psp,N}(R, L_w, L_u, D_1, \mathcal{W}_K) &= \max_{p_{\mathbf{w}} \in \mathcal{P}_W^{[N]}} \min_{p_{\mathbf{s}|\mathbf{w}} \in \mathcal{P}_{S|W}^{[N]}} \max_{p_{\mathbf{xu}|\mathbf{sw}} \in \mathcal{P}_{XU|SW}^{[N]}(p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, L_w, L_u, D_1)} \\ &\quad \hat{E}_{psp,1,N}(R, L_w, L_u, p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_{K_{nom}}^{fair}). \end{aligned} \quad (\text{A.5})$$

Denote by $p_{\mathbf{w}}^*$ and $p_{\mathbf{xu}|\mathbf{sw}}^*$ the maximizers above, the latter viewed as a function of $p_{\mathbf{s}|\mathbf{w}}$. Both maximizers depend implicitly on R and $\mathcal{W}_{K_{nom}}^{fair}$. Let

$$\overline{E}_{psp,N}(R, L_w, L_u, D_1, \mathcal{W}_K) = \min_{p_{\mathbf{s}|\mathbf{w}} \in \mathcal{P}_{S|W}^{[N]}} \overline{\hat{E}}_{psp,N}(R, L_w, L_u, p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}^*), \quad (\text{A.6})$$

$$\underline{E}_{psp,N}(R, L_w, L_u, D_1, \mathcal{W}_K) = \min_{p_{\mathbf{s}|\mathbf{w}} \in \mathcal{P}_{S|W}^{[N]}} \underline{\hat{E}}_{psp,N}(R, L_w, L_u, p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}^*). \quad (\text{A.7})$$

The exponents (A.2)—(A.7) differ from (4.4)—(4.9) in that the optimizations are performed over conditional types instead of general conditional p.m.f.'s. We have

$$\lim_{N \rightarrow \infty} \overline{E}_{p_{sp},N}(R, L_w, L_u, D_1, \mathcal{W}_K) = \overline{E}_{p_{sp}}(R, L_w, L_u, D_1, \mathcal{W}_K) \quad (\text{A.8})$$

$$\lim_{N \rightarrow \infty} \underline{E}_{p_{sp},N}(R, L_w, L_u, D_1, \mathcal{W}_K) = \underline{E}_{p_{sp}}(R, L_w, L_u, D_1, \mathcal{W}_K) \quad (\text{A.9})$$

by continuity of the divergence and mutual-information functionals.

Consider the maximization over the conditional type $p_{\mathbf{x}|\mathbf{u}|\mathbf{s}|\mathbf{w}}$ in (A.5). As a result of this maximization, we may associate the following:

- to any (\mathbf{s}, \mathbf{w}) , a conditional type class $T_{U|SW}^*(\mathbf{s}, \mathbf{w}) \triangleq T_{\mathbf{u}|\mathbf{s}|\mathbf{w}}^*$;
- to any $(\mathbf{s}^d, \mathbf{w})$, a conditional type class $T_{U|S^dW}^*(\mathbf{s}^d, \mathbf{w}) \triangleq T_{\mathbf{u}|\mathbf{s}^d|\mathbf{w}}^*$;
- to any (\mathbf{s}, \mathbf{w}) and $\mathbf{u} \in T_{U|SW}^*(\mathbf{s}, \mathbf{w})$, a conditional type class $T_{X|USW}^*(\mathbf{u}, \mathbf{s}, \mathbf{w}) \triangleq T_{\mathbf{x}|\mathbf{u}|\mathbf{s}|\mathbf{w}}^*$;
- to any type $p_{\mathbf{s}|\mathbf{w}}$, a conditional mutual information $I_{US|S^dW}^*(p_{\mathbf{s}|\mathbf{w}}) \triangleq I(\mathbf{u}; \mathbf{s}|\mathbf{s}^d, \mathbf{w})$ where $\mathbf{u}, \mathbf{s}, \mathbf{w}$ are any three sequences with joint type $p_{\mathbf{u}|\mathbf{s}|\mathbf{w}}^* p_{\mathbf{s}|\mathbf{w}}$.

Codebook. Define the function

$$\rho(p_{\mathbf{s}|\mathbf{s}^d|\mathbf{w}}) = I_{US|S^dW}^*(p_{\mathbf{s}|\mathbf{w}}) + \epsilon, \quad \forall p_{\mathbf{s}|\mathbf{w}} \in \mathcal{P}_{SW}^{[N]}.$$

A random constant-composition code

$$\mathcal{C}(\mathbf{s}^d, \mathbf{w}, p_{\mathbf{s}|\mathbf{s}^d|\mathbf{w}}) = \{\mathbf{u}(l, m, p_{\mathbf{s}|\mathbf{s}^d|\mathbf{w}}), \quad 1 \leq l \leq \exp_2\{N\rho(p_{\mathbf{s}|\mathbf{s}^d|\mathbf{w}})\}, \quad 1 \leq m \leq 2^{NR}\}$$

is generated for each $\mathbf{s}^d \in (\mathcal{S}^d)^N$, $\mathbf{w} \in T_{\mathbf{w}}^*$, and $p_{\mathbf{s}|\mathbf{s}^d|\mathbf{w}} \in \mathcal{P}_{S|S^dW}^{[N]}$ by drawing $\exp_2\{N(R + \rho(p_{\mathbf{s}|\mathbf{s}^d|\mathbf{w}}))\}$ random sequences independently and uniformly from the conditional type class $T_{U|S^dW}^*(\mathbf{s}^d, \mathbf{w})$, and arranging them into an array with 2^{NR} columns and $\exp_2\{N\rho(p_{\mathbf{s}|\mathbf{s}^d|\mathbf{w}})\}$ rows.

Encoder. Prior to encoding, a sequence $\mathbf{W} \in \mathcal{W}^N$ is drawn independently of \mathbf{S} and uniformly from $T_{\mathbf{w}}^*$, and shared with the receiver. Given (\mathbf{S}, \mathbf{W}) , the encoder determines the conditional type $p_{\mathbf{s}|\mathbf{s}^d|\mathbf{w}}$ and performs the following two steps for each user $1 \leq m \leq 2^{NR}$.

- 1) Find l such that $\mathbf{u}(l, m, p_{\mathbf{s}|\mathbf{s}^d|\mathbf{w}}) \in \mathcal{C}(\mathbf{S}^d, \mathbf{W}, p_{\mathbf{s}|\mathbf{s}^d|\mathbf{w}}) \cap T_{U|SW}^*(\mathbf{s}, \mathbf{w})$. If more than one such l exists, pick one of them randomly (with uniform distribution). Let $\mathbf{u} = \mathbf{u}(l, m, p_{\mathbf{s}|\mathbf{s}^d|\mathbf{w}})$. If no such l can be found, generate \mathbf{u} uniformly from the conditional type class $T_{U|SW}^*(\mathbf{s}, \mathbf{w})$.
- 2) Generate \mathbf{X}_m uniformly distributed over the conditional type class $T_{X|USW}^*(\mathbf{u}, \mathbf{s}, \mathbf{w})$.

Collusion channel. By Prop. 3.1, it is sufficient to restrict our attention to strongly exchangeable collusion channels in the error probability analysis.

Decoder. Given $(\mathbf{y}, \mathbf{s}^d, \mathbf{w})$, the decoder outputs $\hat{\mathcal{K}}$ if and only if (4.1) is satisfied.

Encoding errors. Analogously to [11], the probability of encoding errors vanishes doubly exponentially with N because $\rho(p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}) > I(\mathbf{u}; \mathbf{s}|\mathbf{s}^d\mathbf{w})$. Indeed an encoding error for user m arises under the following event:

$$\mathcal{E}_m = \{(\mathcal{C}, \mathbf{s}, \mathbf{w}) : (\mathbf{u}(l, m, p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}) \in \mathcal{C} \text{ and } \mathbf{u}(l, m, p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}) \notin T_{U|SW}^*(\mathbf{s}, \mathbf{w})) \text{ for } 1 \leq l \leq 2^{N\rho(p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}})}\}. \quad (\text{A.10})$$

The probability that a sequence \mathbf{U} uniformly distributed over $T_{U|S^dW}^*(\mathbf{s}^d, \mathbf{w})$ also belongs to $T_{U|SW}^*(\mathbf{s}, \mathbf{w})$ is equal to $\exp_2\{-NI_{US|S^dW}^*(p_{\mathbf{sw}})\}$ on the exponential scale. Therefore the encoding error probability, conditioned on type class $T_{\mathbf{sw}}$, satisfies

$$\begin{aligned} Pr[\mathcal{E}_m | (\mathbf{S}, \mathbf{W}) \in T_{\mathbf{sw}}] &= \left(1 - \frac{|T_{U|SW}^*(\mathbf{S}, \mathbf{W})|}{|T_{U|S^dW}^*(\mathbf{S}^d, \mathbf{W})|}\right)^{2^{N\rho(p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}})}} \\ &\doteq (1 - 2^{-NI_{US|S^dW}^*(p_{\mathbf{sw}})})^{2^{N\rho(p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}})}} \\ &\leq \exp\{-\exp_2(N[\rho(p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}) - I_{US|S^dW}^*(p_{\mathbf{sw}})])\} \\ &= \exp\{-2^{N\epsilon}\} \end{aligned} \quad (\text{A.11})$$

where the inequality follows from $1 - a \leq e^{-a}$.

The derivation of the decoding error exponents is based on the following two asymptotic equalities which are special cases of (C.2) and (C.5) established in Lemma 3.1.

1) Fix $\mathbf{y}, \mathbf{s}^d, \mathbf{w}$ and draw \mathbf{u} uniformly from some fixed type class, independently of $(\mathbf{y}, \mathbf{s}^d, \mathbf{w})$. Then

$$Pr[I(\mathbf{u}; \mathbf{y}|\mathbf{s}^d\mathbf{w}) \geq \nu] \doteq 2^{-N\nu}. \quad (\text{A.12})$$

2) Given \mathbf{s}, \mathbf{w} , draw $(\mathbf{x}_k, \mathbf{u}_k)$, $k \in \mathcal{K}$, i.i.d. uniformly from a conditional type class $T_{\mathbf{xu}|\mathbf{sw}}$, and then draw \mathbf{y} uniformly over a single conditional type class $T_{\mathbf{y}|\mathbf{x}_\mathcal{K}}$. For any $\nu > 0$, we have

$$Pr[I(\mathbf{u}_m; \mathbf{y}|\mathbf{s}^d\mathbf{w}) \leq \rho(p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}) + \nu] \doteq \exp_2\{-N\check{E}_{psp,m,N}(\nu, L_w, L_u, p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}^*, \mathcal{W}_K)\}. \quad (\text{A.13})$$

(i). False Positives. From (4.1), the occurrence of a false positive implies that

$$\exists \lambda \in \mathcal{S}_{S|S^dW}^{[N]}, l, m \notin \mathcal{K} : I(\mathbf{u}(l, m, \lambda); \mathbf{y}|\mathbf{s}^d\mathbf{w}) > \rho(\lambda) + R + \Delta. \quad (\text{A.14})$$

By construction of the codebook, $\mathbf{u}(l, m, \lambda)$ is independent of \mathbf{y} for $m \notin \mathcal{K}$. For any given λ , there are at most $2^{N\rho(\lambda)}$ possible values for l and $2^{NR} - K$ possible values for m in (A.14). Hence the probability

of false positives, conditioned on the joint type class $T_{\mathbf{y}(\mathbf{xu})_{\mathcal{K}}\mathbf{sw}}$, is

$$\begin{aligned}
& P_{FP}(T_{\mathbf{y}(\mathbf{xu})_{\mathcal{K}}\mathbf{sw}}, \mathcal{W}_K) \\
& \leq \sum_{\lambda} (2^{NR} - K) 2^{N\rho(\lambda)} Pr[I(\mathbf{u}(l, m, \lambda); \mathbf{y}|\mathbf{s}^d\mathbf{w}) > \rho(\lambda) + R + \Delta] \\
& \stackrel{(a)}{=} \sum_{\lambda} 2^{N(R+\rho(\lambda))} 2^{-N(R+\Delta+\rho(\lambda))} \\
& \stackrel{(b)}{\leq} (N+1)^{|S|L_w} 2^{-N\Delta} \\
& \doteq 2^{-N\Delta}
\end{aligned} \tag{A.15}$$

where (a) is obtained by application of (A.12) with $\nu = \rho(\lambda) + R + \Delta$, and (b) because the number of conditional types λ is at most $(N+1)^{|S|L_w}$.

Averaging over all type classes $T_{\mathbf{y}(\mathbf{xu})_{\mathcal{K}}\mathbf{sw}}$, we obtain $P_{FP} \stackrel{\cdot}{\leq} 2^{-N\Delta}$, from which (4.10) follows.

(ii). Detect-One Error Criterion (Miss All Colluders). We first derive the error exponent for the event that the decoder misses a specific colluder $m \in \mathcal{K}$. Any coalition $\hat{\mathcal{K}}$ that contains m fails the test (4.1), i.e., for any such $\hat{\mathcal{K}}$,

$$\forall \lambda \in \mathcal{P}_{S|S^dW}^{[N]} : \quad \max_l I(\mathbf{u}(l, m, \lambda); \mathbf{y}|\mathbf{s}^d\mathbf{w}) \leq \rho(\lambda) + R + \Delta. \tag{A.16}$$

This implies that

$$I(\mathbf{u}(l, m, p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}); \mathbf{y}|\mathbf{s}^d\mathbf{w}) \leq \rho(p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}) + R + \Delta \tag{A.17}$$

where l is the row index actually selected by the encoder, and $p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}$ is the actual host sequence conditional type. The probability of the miss- m event, given the joint type $p_{\mathbf{w}}^* p_{\mathbf{s}|\mathbf{w}} p_{\mathbf{xu}|\mathbf{sw}}^*$, is therefore upper-bounded by the probability of the event (A.17):

$$\begin{aligned}
p_{miss-m}(p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}^*, \mathcal{W}_K) & \leq Pr \left[I(\mathbf{u}(l, m, p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}); \mathbf{y}|\mathbf{s}^d\mathbf{w}) \leq \rho(p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}) + R + \Delta \right] \\
& \stackrel{(a)}{\leq} \exp_2 \left\{ -N \check{E}_{psp,m,N}(R + \Delta, L_w, L_u, p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}^*, \mathcal{W}_K) \right\}
\end{aligned}$$

where (a) follows from (A.13) with $\nu = R + \Delta$.

The miss-all event is the intersection of the miss- m events over $m \in \mathcal{K}$. Its conditional probability is

$$\begin{aligned}
& p_{\text{miss-all}}(p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}^*, \mathcal{W}_K) \\
&= Pr \left[\bigcap_{m \in \mathcal{K}} \left\{ \text{miss } m \mid p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}^* \right\} \right] \\
&\leq \min_{m \in \mathcal{K}} p_{\text{miss}-m}(p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}^*, \mathcal{W}_K) \\
&\doteq \exp_2 \left\{ -N \max_{m \in \mathcal{K}} \check{E}_{psp,m,N}(R + \Delta, L_w, L_u, p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}^*, \mathcal{W}_K) \right\}. \tag{A.18}
\end{aligned}$$

Averaging over \mathbf{S} , we obtain

$$\begin{aligned}
& p_{\text{miss-all}}(\mathcal{W}_K) \\
&\leq \sum_{p_{\mathbf{s}|\mathbf{w}}} Pr[T_{\mathbf{s}|\mathbf{w}}] p_{\text{miss-all}}(p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}^*, \mathcal{W}_K) \\
&\doteq \max_{p_{\mathbf{s}|\mathbf{w}}} Pr[T_{\mathbf{s}|\mathbf{w}}] p_{\text{miss-all}}(p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}^*, \mathcal{W}_K) \\
&\stackrel{(a)}{\doteq} \max_{p_{\mathbf{s}|\mathbf{w}}} \exp_2 \left\{ -N \left[D(p_{\mathbf{s}|\mathbf{w}} \| p_{\mathbf{S}} | p_{\mathbf{w}}^*) + \max_{m \in \mathcal{K}} \check{E}_{psp,m,N}(R + \Delta, L_w, L_u, p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}^*, \mathcal{W}_K) \right] \right\} \\
&\stackrel{(b)}{\doteq} \exp_2 \left\{ -N \overline{E}_{psp,N}(R + \Delta, L_w, L_u, D_1, \mathcal{W}_K) \right\} \\
&\stackrel{(c)}{\doteq} \exp_2 \left\{ -N \overline{E}_{psp}(R + \Delta, L_w, L_u, D_1, \mathcal{W}_K) \right\}
\end{aligned}$$

which establishes (4.12). Here (a) follows from (C.3) and (A.18), (b) from (A.3) and (A.6), and (c) from (A.8).

(iii). Detect-All Error Criterion (Miss Some Colluders).

The miss-some event is the union of the miss- m events over $m \in \mathcal{K}$. Given the joint type $p_{\mathbf{w}}^* p_{\mathbf{s}|\mathbf{w}} p_{\mathbf{xu}|\mathbf{sw}}^*$, the probability of this event is

$$\begin{aligned}
& p_{\text{miss-some}}(p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}^*, \mathcal{W}_K) \tag{A.19} \\
&= Pr \left[\bigcup_{m \in \mathcal{K}} \left\{ \text{miss } m \mid p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}^* \right\} \right] \\
&\leq \sum_{m \in \mathcal{K}} p_{\text{miss}-m}(p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}^*, \mathcal{W}_K) \\
&\doteq \max_{m \in \mathcal{K}} \exp_2 \left\{ -N \check{E}_{psp,m,N}(R + \Delta, L_w, L_u, p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}^*, \mathcal{W}_K) \right\} \\
&= \exp_2 \left\{ -N \min_{m \in \mathcal{K}} \check{E}_{psp,m,N}(R + \Delta, L_w, L_u, p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}^*, \mathcal{W}_K) \right\}. \tag{A.20}
\end{aligned}$$

Averaging over \mathbf{S} , we obtain

$$\begin{aligned}
& p_{\text{miss-some}}(\mathcal{W}_K) \\
& \leq \sum_{p_{\mathbf{S}|\mathbf{W}}} Pr[T_{\mathbf{S}|\mathbf{W}}] p_{\text{miss-some}}(p_{\mathbf{W}}^*, p_{\mathbf{S}|\mathbf{W}}, p_{\mathbf{XU}|\mathbf{SW}}^*, \mathcal{W}_K) \\
& \stackrel{(a)}{=} \max_{p_{\mathbf{S}|\mathbf{W}}} \exp_2 \left\{ -N \left[D(p_{\mathbf{S}|\mathbf{W}} \| p_{\mathbf{S}} | p_{\mathbf{W}}^*) + \min_{m \in \mathcal{K}} \check{E}_{psp,m,N}(R + \Delta, L_w, L_u, p_{\mathbf{W}}^*, p_{\mathbf{S}|\mathbf{W}}, p_{\mathbf{XU}|\mathbf{SW}}^*, \mathcal{W}_K) \right] \right\} \\
& \stackrel{(b)}{\leq} \exp_2 \left\{ -N \underline{E}_{psp,N}(R + \Delta, D_1, \mathcal{W}_K) \right\} \\
& \stackrel{(c)}{=} \exp_2 \left\{ -N \underline{E}_{psp}(R + \Delta, L_w, L_u, D_1, \mathcal{W}_K) \right\}
\end{aligned}$$

which establishes (4.11). Here (a) follows from (C.3) and (A.20), (b) from (A.7) and (A.4), and (c) from (A.9).

(iv). Fair Collusion Channels. The proof parallels that of [10, Theorem 4.1(iv)], using the conditional divergence $D(\tilde{p}_{Y(XU)_{\mathcal{K}}|SW} \tilde{p}_{S|W} \| \tilde{p}_{Y|X_{\mathcal{K}}} p_{XU|SW}^{\mathcal{K}} p_{\mathbf{S}} | p_W)$ in place of $D(\tilde{p}_{YX_{\mathcal{K}}|W} \| \tilde{p}_{Y|X_{\mathcal{K}}} p_{X|W}^{\mathcal{K}} | p_W)$.

(v). Immediate, because $\bar{E}_{psp} = \underline{E}_{psp}$ in this case.

(vi). Positive Error Exponents. From Part (v) above, we may restrict our attention to $\mathcal{W}_K = \mathcal{W}_K^{\text{fair}}$. Consider any $\mathcal{W} = \{1, \dots, L_w\}$ and p_W that is positive over its support set (if it is not, reduce the value of L_w accordingly.) For any $m \in \mathcal{K}$, the minimand in the expression (4.4) for $\tilde{E}_{psp,m}(R, L_w, L_u, p_W, p_{XU|SW}, \mathcal{W}_K^{\text{fair}})$ is zero if and only if

$$\tilde{p}_{Y(XU)_{\mathcal{K}}|SW} \tilde{p}_{S|W} = \tilde{p}_{Y|X_{\mathcal{K}}} p_{XU|SW}^{\mathcal{K}} p_{\mathbf{S}}, \quad \text{with } \tilde{p}_{Y|X_{\mathcal{K}}} \in \mathcal{W}_K^{\text{fair}}.$$

Such $(\tilde{p}_{Y(XU)_{\mathcal{K}}|SW}, \tilde{p}_{S|W})$ is feasible for (4.3) if and only if $(p_{XU|SW}, \tilde{p}_{Y|X_{\mathcal{K}}})$ is such that $I(U_m; Y | S^d, W) \leq I(U_m; S | S^d, W) + R$. It is not feasible, and thus a positive exponent E^{one} is guaranteed, if $R < I(U_1; Y | S^d, W) - I(U_1; S | S^d, W)$. The supremum of all such R is given by (4.13) and is achieved by letting $\epsilon \rightarrow 0$, $\Delta \rightarrow 0$, and $L_w, L_u \rightarrow \infty$. \square

APPENDIX II

PROOF OF THEOREM 5.2

We derive the error exponents for the M2PMI decision rule (5.2). Define for all $\mathcal{A} \subseteq \mathcal{K}$

$$\begin{aligned} \mathcal{P}_{Y(XU)\mathcal{K}|SW}^{[N]}(p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_K, R, L_w, L_u, \mathcal{A}) &= \left\{ p_{\mathbf{y}(\mathbf{xu})_{\mathcal{K}}|\mathbf{sw}} : p_{(\mathbf{xu})_{\mathcal{K}}|\mathbf{sw}} \in \mathcal{M}(p_{\mathbf{xu}|\mathbf{sw}}), \right. \\ &\quad p_{\mathbf{y}|\mathbf{x}_{\mathcal{K}}} \in \mathcal{W}_K, \\ &\quad \left. \mathring{I}(\mathbf{u}_{\mathcal{A}}; \mathbf{y}_{\mathbf{u}_{\mathcal{K} \setminus \mathcal{A}}} | \mathbf{s}^d \mathbf{w}) \leq |\mathcal{A}|(\rho(p_{\mathbf{s}|\mathbf{s}^d \mathbf{w}}) + R) \right\} \quad (\text{B.1}) \end{aligned}$$

$$\begin{aligned} \check{E}_{psp, \mathcal{A}, N}(R, L_w, L_u, p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_K) &= \min_{p_{\mathbf{y}(\mathbf{xu})_{\mathcal{K}}|\mathbf{sw}} \in \mathcal{P}_{Y(XU)\mathcal{K}|SW}^{[N]}(p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_K, R, L_w, L_u, \mathcal{A})} \\ &\quad D(p_{\mathbf{y}(\mathbf{xu})_{\mathcal{K}}|\mathbf{sw}} \| p_{\mathbf{y}|\mathbf{x}_{\mathcal{K}}} p_{\mathbf{xu}|\mathbf{sw}}^{\mathcal{K}} | p_{\mathbf{sw}}), \quad (\text{B.2}) \end{aligned}$$

$$\begin{aligned} \hat{E}_{psp, \mathcal{A}, N}(R, L_w, L_u, p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_K) &= D(p_{\mathbf{s}|\mathbf{w}} \| p_{\mathbf{S}} | p_{\mathbf{w}}) + \check{E}_{psp, \mathcal{A}, N}(R, L_w, L_u, \\ &\quad p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_K) \\ &= \min_{p_{\mathbf{y}(\mathbf{xu})_{\mathcal{K}}|\mathbf{sw}} \in \mathcal{P}_{Y(XU)\mathcal{K}|SW}^{[N]}(p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_K, R, L_w, L_u, \mathcal{A})} \\ &\quad D(p_{\mathbf{y}(\mathbf{xu})_{\mathcal{K}}|\mathbf{sw}} p_{\mathbf{s}|\mathbf{w}} \| p_{\mathbf{y}|\mathbf{x}_{\mathcal{K}}} p_{\mathbf{xu}|\mathbf{sw}}^{\mathcal{K}} p_{\mathbf{S}} | p_{\mathbf{w}}), \quad (\text{B.3}) \end{aligned}$$

$$\overline{\hat{E}}_{psp, N}(R, L_w, L_u, p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_K) = \hat{E}_{psp, \mathcal{K}, N}(R, L_w, L_u, p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_K), \quad (\text{B.4})$$

$$\begin{aligned} \underline{\hat{E}}_{psp, N}(R, L_w, L_u, p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_K) &= \min_{\mathcal{A} \subseteq \mathcal{K}} \hat{E}_{psp, \mathcal{A}, N}(R, L_w, L_u, p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_K), \\ &\quad (\text{B.5}) \end{aligned}$$

$$\begin{aligned} E_{psp, N}(R, L_w, L_u, D_1, \mathcal{W}_K) &= \max_{p_{\mathbf{w}} \in \mathcal{P}_W^{[N]}} \min_{p_{\mathbf{s}|\mathbf{w}} \in \mathcal{P}_{S|W}^{[N]} \quad p_{\mathbf{xu}|\mathbf{sw}} \in \mathcal{P}_{XU|SW}^{[N]}(p_{\mathbf{w}} p_{\mathbf{s}|\mathbf{w}}, L_w, L_u, D_1)} \max \\ &\quad \hat{E}_{psp, \mathcal{K}, N}(R, L_w, L_u, p_{\mathbf{w}}, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}, \mathcal{W}_{K_{nom}}^{fair}). \quad (\text{B.6}) \end{aligned}$$

Denote by $p_{\mathbf{w}}^*$ and $p_{\mathbf{xu}|\mathbf{sw}}^*$ the maximizers in (B.6), the latter viewed as a function of $p_{\mathbf{s}|\mathbf{w}}$. Both maximizers depend implicitly on R , D_1 , and $\mathcal{W}_{K_{nom}}^{fair}$. Let

$$\overline{E}_{psp, N}(R, L_w, L_u, D_1, \mathcal{W}_K) = \min_{p_{\mathbf{s}|\mathbf{w}}} \overline{\hat{E}}_{psp, N}(R, L_w, L_u, p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}^*, \mathcal{W}_K) \quad (\text{B.7})$$

$$\underline{E}_{psp, N}(R, L_w, L_u, D_1, \mathcal{W}_K) = \min_{p_{\mathbf{s}|\mathbf{w}}} \underline{\hat{E}}_{psp, N}(R, L_w, L_u, p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}^*, \mathcal{W}_K). \quad (\text{B.8})$$

The exponents (B.3)—(B.8) differ from (5.7)—(5.12) in that the optimizations are performed over

conditional types instead of general conditional p.m.f.'s. We have

$$\lim_{N \rightarrow \infty} \overline{E}_{p_{sp}, N}(R, L_w, L_u, D_1, \mathcal{W}_K) = \overline{E}_{p_{sp}}(R, L_w, L_u, D_1, \mathcal{W}_K) \quad (\text{B.9})$$

$$\lim_{N \rightarrow \infty} \underline{E}_{p_{sp}, N}(R, L_w, L_u, D_1, \mathcal{W}_K) = \underline{E}_{p_{sp}}(R, L_w, L_u, D_1, \mathcal{W}_K) \quad (\text{B.10})$$

by continuity of the divergence and mutual-information functionals.

The codebook and encoding procedure are exactly as in the proof of Theorem IV, the difference being that $p_{\mathbf{w}}^*$ and $p_{\mathbf{xu}|\mathbf{sw}}^*$ are solutions to the optimization problem (B.6) instead of (A.5). The decoding rule is the M2PMI rule of (5.2).

To analyze the error probability for this random-coding scheme, it is again sufficient to restrict our attention to strongly-exchangeable channels and use the bound (3.2) on the conditional probability of the collusion channel output. We also use Lemma 3.1.

(i). False Positives. By application of (5.4), a false positive occurs if $\hat{\mathcal{K}} \setminus \mathcal{K} \neq \emptyset$ and

$$\begin{aligned} \exists \lambda \in \mathcal{P}_{S|S^d \mathbf{W}}^{[N]} : \quad \forall \mathcal{A} \subseteq \hat{\mathcal{K}} : \quad \exists l_{\hat{\mathcal{K}}} : \quad & \overset{\circ}{I}(\mathbf{u}(l_{\mathcal{A}}, m_{\mathcal{A}}, \lambda); \mathbf{y}\mathbf{u}(l_{\hat{\mathcal{K}} \setminus \mathcal{A}}, m_{\hat{\mathcal{K}} \setminus \mathcal{A}}, \lambda) | \mathbf{s}^d \mathbf{w}) \\ & > |\mathcal{A}| (\rho(\lambda) + R + \Delta). \end{aligned} \quad (\text{B.11})$$

The probability of this event is upper-bounded by the probability of the larger event

$$\begin{aligned} \forall \mathcal{A} \subseteq \hat{\mathcal{K}} : \quad \exists \lambda, l_{\hat{\mathcal{K}}} : \quad & \overset{\circ}{I}(\mathbf{u}(l_{\mathcal{A}}, m_{\mathcal{A}}, \lambda); \mathbf{y}\mathbf{u}(l_{\hat{\mathcal{K}} \setminus \mathcal{A}}, m_{\hat{\mathcal{K}} \setminus \mathcal{A}}, \lambda) | \mathbf{s}^d \mathbf{w}) \\ & > |\mathcal{A}| (\rho(\lambda) + R + \Delta). \end{aligned} \quad (\text{B.12})$$

Denote by $p_{\mathbf{s}|\mathbf{s}^d \mathbf{w}}^*$ the conditional type of the host sequence and by $l_{\mathcal{K}}^*$ the row indices selected by the encoder. To each triple $(\hat{\mathcal{K}}, \lambda, l_{\hat{\mathcal{K}}})$, we associate a unique subset \mathcal{B} of $\mathcal{K} \cap \hat{\mathcal{K}}$ defined as follows:

- If $\lambda \neq p_{\mathbf{s}|\mathbf{s}^d \mathbf{w}}^*$ then $\mathcal{B} = \emptyset$
- If $\lambda = p_{\mathbf{s}|\mathbf{s}^d \mathbf{w}}^*$ then \mathcal{B} is the (possibly empty) set of all indices $k \in \mathcal{K} \cap \hat{\mathcal{K}}$ such that $l_k = l_k^*$.

Thus \mathcal{B} is the set of colluder indices $k \in \mathcal{K}$ for which the decoder correctly identifies the conditional host sequence type $p_{\mathbf{s}|\mathbf{s}^d \mathbf{w}}^*$ and the codewords $\mathbf{u}(l_k^*, k, p_{\mathbf{s}|\mathbf{s}^d \mathbf{w}}^*)$ that were assigned by the encoder. Denoting by $\Omega(\mathcal{B})$ the set of pairs $(\lambda, l_{\hat{\mathcal{K}}})$ associated with \mathcal{B} , we rewrite (B.12) as

$$\begin{aligned} \forall \mathcal{A} \subseteq \hat{\mathcal{K}} : \quad \exists \mathcal{B} \subseteq \mathcal{K} \cap \hat{\mathcal{K}}, \exists (\lambda, l_{\hat{\mathcal{K}}}) \in \Omega(\mathcal{B}) : \\ \overset{\circ}{I}(\mathbf{u}(l_{\mathcal{A}}, m_{\mathcal{A}}, \lambda); \mathbf{y}\mathbf{u}(l_{\hat{\mathcal{K}} \setminus \mathcal{A}}, m_{\hat{\mathcal{K}} \setminus \mathcal{A}}, \lambda) | \mathbf{s}^d \mathbf{w}) > |\mathcal{A}| (\rho(\lambda) + R + \Delta). \end{aligned} \quad (\text{B.13})$$

Define the complement set $\mathcal{A} = \hat{\mathcal{K}} \setminus \mathcal{B}$ which is comprised of all incorrectly accused users as well as any colluder k such that $\lambda \neq p_{\mathbf{s}|\mathbf{s}^d \mathbf{w}}^*$ or $l_k \neq l_k^*$. Since $\mathcal{B} \subseteq \hat{\mathcal{K}}$ and there is at least one innocent user in $\hat{\mathcal{K}}$, the cardinality of \mathcal{A} is at least equal to 1. By construction of the codebook and definition of \mathcal{A}

and \mathcal{B} , $\mathbf{u}(l_{\mathcal{A}}, m_{\mathcal{A}}, \lambda)$ is independent of \mathbf{y} and $\mathbf{u}(l_{\mathcal{B}}^*, m_{\mathcal{B}}, p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}^*)$. The probability of the event (B.13) is upper-bounded by the probability of the larger event

$$\exists \mathcal{B} \subseteq \mathcal{K}, \exists \lambda, l_{\mathcal{A}}, m_{\mathcal{A}} : \quad \overset{\circ}{I}(\mathbf{u}(l_{\mathcal{A}}, m_{\mathcal{A}}, \lambda); \mathbf{y}\mathbf{u}(l_{\mathcal{B}}^*, m_{\mathcal{B}}, p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}^*) | \mathbf{s}^d\mathbf{w}) > |\mathcal{A}|(\rho(\lambda) + R + \Delta). \quad (\text{B.14})$$

Hence the probability of false positives, conditioned on $T_{\mathbf{y}(\mathbf{xu})_{\mathcal{K}\mathbf{sw}}}$, satisfies

$$\begin{aligned} & P_{FP}(T_{\mathbf{y}(\mathbf{xu})_{\mathcal{K}\mathbf{sw}}}, \mathscr{W}_K) \\ &= Pr \left[\bigcup_{\mathcal{B} \subseteq \mathcal{K}} \bigcup_{|\mathcal{A}| \geq 1} \left\{ \exists \lambda, l_{\mathcal{A}}, m_{\mathcal{A}} : \overset{\circ}{I}(\mathbf{u}(l_{\mathcal{A}}, m_{\mathcal{A}}, \lambda); \mathbf{y}\mathbf{u}(l_{\mathcal{B}}^*, m_{\mathcal{B}}, p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}^*) | \mathbf{s}^d\mathbf{w}) \right. \right. \\ &\quad \left. \left. > |\mathcal{A}|(\rho(\lambda) + R + \Delta) \right\} \right] \\ &\leq \sum_{\mathcal{B} \subseteq \mathcal{K}} \sum_{|\mathcal{A}| \geq 1} P_{\mathcal{B}, |\mathcal{A}|}(T_{\mathbf{y}(\mathbf{xu})_{\mathcal{K}\mathbf{sw}}}, \mathscr{W}_K) \end{aligned} \quad (\text{B.15})$$

where

$$\begin{aligned} P_{\mathcal{B}, |\mathcal{A}|}(T_{\mathbf{y}(\mathbf{xu})_{\mathcal{K}\mathbf{sw}}}, \mathscr{W}_K) &= Pr [\exists \lambda, l_{\mathcal{A}}, m_{\mathcal{A}} : \quad \overset{\circ}{I}(\mathbf{u}(l_{\mathcal{A}}, m_{\mathcal{A}}, \lambda); \mathbf{y}\mathbf{u}(l_{\mathcal{B}}^*, m_{\mathcal{B}}, p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}^*) | \mathbf{s}^d\mathbf{w}) \\ &\quad > |\mathcal{A}|(\rho(\lambda) + R + \Delta)]. \end{aligned} \quad (\text{B.16})$$

By definition of \mathcal{B} , there are at most $\sum_{\lambda \neq p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}} 2^{N|\mathcal{A}|\rho(\lambda)}$ possible values for $l_{\mathcal{A}}$ and $2^{N|\mathcal{A}|R}$ possible values for $m_{\mathcal{A}}$ in (B.16). Hence

$$\begin{aligned} & P_{\mathcal{B}, |\mathcal{A}|}(T_{\mathbf{y}(\mathbf{xu})_{\mathcal{K}\mathbf{sw}}}, \mathscr{W}_K) \\ &\leq \sum_{\lambda} 2^{N|\mathcal{A}|(R+\rho(\lambda))} Pr[\overset{\circ}{I}(\mathbf{u}(l_{\mathcal{A}}, m_{\mathcal{A}}, \lambda); \mathbf{y}\mathbf{u}(l_{\mathcal{B}}^*, m_{\mathcal{B}}, p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}^*) | \mathbf{s}^d\mathbf{w}) > |\mathcal{A}|(\rho(\lambda) + R + \Delta)] \\ &\stackrel{(a)}{=} \sum_{\lambda} 2^{N|\mathcal{A}|(R+\rho(\lambda))} 2^{-N|\mathcal{A}|(R+\Delta+\rho(\lambda))} \\ &\leq (N+1)^{|\mathcal{S}|} 2^{-N|\mathcal{A}|\Delta} \\ &\doteq 2^{-N|\mathcal{A}|\Delta} \end{aligned} \quad (\text{B.17})$$

where (a) is obtained by application of (C.2) with $\mathbf{y}\mathbf{u}(l_{\mathcal{B}}^*, m_{\mathcal{B}}, p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}^*)$ in place of \mathbf{z} .

Combining (B.15) and (B.17) we obtain

$$\begin{aligned} P_{FP}(T_{\mathbf{y}(\mathbf{xu})_{\mathcal{K}\mathbf{sw}}}, \mathscr{W}_K) &\leq \sum_{\mathcal{B} \subseteq \mathcal{K}} \sum_{|\mathcal{A}| \geq 1} 2^{-N|\mathcal{A}|\Delta} \\ &\doteq 2^{-N\Delta}. \end{aligned}$$

Averaging over all joint type classes $T_{\mathbf{y}(\mathbf{xu})_{\mathcal{K}\mathbf{sw}}}$, we obtain $P_{FP} \leq 2^{-N\Delta}$, from which (5.13) follows.

(ii). Detect-All Criterion. (Miss Some Colluders.)

Under the miss-some error event, *any* coalition $\hat{\mathcal{K}}$ that *contains* \mathcal{K} fails the test. By (5.4), this implies

$$\begin{aligned} \forall \lambda \in \mathcal{P}_{S|S^dW}^{[N]} : \quad \exists \mathcal{A} \subseteq \hat{\mathcal{K}} : \quad \max_{l_{\hat{\mathcal{K}}}} \overset{\circ}{I}(\mathbf{u}(l_{\mathcal{A}}, m_{\mathcal{A}}, \lambda); \mathbf{y}\mathbf{u}(l_{\hat{\mathcal{K}} \setminus \mathcal{A}}, m_{\hat{\mathcal{K}} \setminus \mathcal{A}}, \lambda) | \mathbf{s}^d \mathbf{w}) \\ \leq |\mathcal{A}| (\rho(\lambda) + R + \Delta). \end{aligned} \quad (\text{B.18})$$

In particular, for $\hat{\mathcal{K}} = \mathcal{K}$ we have

$$\exists \mathcal{A} \subseteq \mathcal{K} : \quad \overset{\circ}{I}(\mathbf{u}(l_{\mathcal{A}}, m_{\mathcal{A}}, p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}); \mathbf{y}\mathbf{u}(l_{\mathcal{K} \setminus \mathcal{A}}, m_{\mathcal{K} \setminus \mathcal{A}}, p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}) | \mathbf{s}^d \mathbf{w}) \leq |\mathcal{A}| (\rho(p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}) + R + \Delta). \quad (\text{B.19})$$

where $l_{\mathcal{K}}$ are the row indices actually selected by the encoder, and $p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}$ is the actual host sequence conditional type. The probability of the miss-some event, conditioned on (\mathbf{s}, \mathbf{w}) , is therefore upper bounded by the probability of the event (B.19):

$$\begin{aligned} p_{\text{miss-some}}(p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{x}\mathbf{u}|\mathbf{s}\mathbf{w}}^*, \mathcal{W}_K) \\ \leq Pr \left[\bigcup_{\mathcal{A} \subseteq \mathcal{K}} \left\{ \overset{\circ}{I}(\mathbf{u}(l_{\mathcal{A}}, m_{\mathcal{A}}, p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}); \mathbf{y}\mathbf{u}(l_{\mathcal{K} \setminus \mathcal{A}}, m_{\mathcal{K} \setminus \mathcal{A}}, p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}) | \mathbf{s}^d \mathbf{w}) \leq |\mathcal{A}| (\rho(p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}) + R + \Delta) \right\} \right] \\ \leq \sum_{\mathcal{A} \subseteq \mathcal{K}} Pr \left[\overset{\circ}{I}(\mathbf{u}(l_{\mathcal{A}}, m_{\mathcal{A}}, p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}); \mathbf{y}\mathbf{u}(l_{\mathcal{K} \setminus \mathcal{A}}, m_{\mathcal{K} \setminus \mathcal{A}}, p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}) | \mathbf{s}^d \mathbf{w}) \leq |\mathcal{A}| (\rho(p_{\mathbf{s}|\mathbf{s}^d\mathbf{w}}) + R + \Delta) \right] \\ \stackrel{(a)}{\leq} \sum_{\mathcal{A} \subseteq \mathcal{K}} \exp_2 \left\{ -N \check{E}_{psp, \mathcal{A}, N}(R + \Delta, L_w, L_u, p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{x}\mathbf{u}|\mathbf{s}\mathbf{w}}^*, \mathcal{W}_K) \right\} \\ \doteq \max_{\mathcal{A} \subseteq \mathcal{K}} \exp_2 \left\{ -N \check{E}_{psp, \mathcal{A}, N}(R + \Delta, L_w, L_u, p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{x}\mathbf{u}|\mathbf{s}\mathbf{w}}^*, \mathcal{W}_K) \right\} \\ = \exp_2 \left\{ -N \min_{\mathcal{A} \subseteq \mathcal{K}} \check{E}_{psp, \mathcal{A}, N}(R + \Delta, L_w, L_u, p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{x}\mathbf{u}|\mathbf{s}\mathbf{w}}^*, \mathcal{W}_K) \right\} \end{aligned} \quad (\text{B.20})$$

where (a) follows from (C.5) with $\nu = R + \Delta$.

Averaging over \mathbf{S} , we obtain

$$\begin{aligned} p_{\text{miss-some}}(\mathcal{W}_K) \\ = \sum_{p_{\mathbf{s}|\mathbf{w}}} Pr[T_{\mathbf{s}|\mathbf{w}}] p_{\text{miss-some}}(p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{x}\mathbf{u}|\mathbf{s}\mathbf{w}}^*, \mathcal{W}_K) \\ \stackrel{(a)}{=} \max_{p_{\mathbf{s}|\mathbf{w}}} \exp_2 \left\{ -N [D(p_{\mathbf{s}|\mathbf{w}} \| p_{\mathbf{S}} | p_{\mathbf{w}}) + \min_{\mathcal{A} \subseteq \mathcal{K}} \check{E}_{psp, N}(R + \Delta, L, p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{x}\mathbf{u}|\mathbf{s}\mathbf{w}}^*, \mathcal{W}_K)] \right\} \\ \stackrel{(b)}{=} \max_{p_{\mathbf{s}|\mathbf{w}}} \exp_2 \left\{ -N \hat{E}_{psp, N}(R + \Delta, L_w, L_u, p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{x}\mathbf{u}|\mathbf{s}\mathbf{w}}^*, \mathcal{W}_K) \right\} \\ \stackrel{(c)}{=} \exp_2 \left\{ -N \underline{E}_{psp, N}(R + \Delta, L_w, L_u, D_1, \mathcal{W}_K) \right\} \\ \stackrel{(d)}{=} \exp_2 \left\{ -N \underline{E}_{psp}(R + \Delta, L_w, L_u, D_1, \mathcal{W}_K) \right\} \end{aligned}$$

which proves (5.14). Here (a) follows from (C.3) and (B.20), (b) from the definitions (B.5) and (B.3), (c) from (B.8), and (d) from the limit property (B.10).

(iii). Detect-One Criterion (Miss All Colluders.) Either the estimated coalition $\hat{\mathcal{K}}$ is empty, or it is a set \mathcal{I} of innocent users (disjoint with \mathcal{K}). Hence $P_e^{one} \leq Pr[\hat{\mathcal{K}} = \emptyset] + Pr[\hat{\mathcal{K}} = \mathcal{I}]$. The first probability, conditioned on $(\mathbf{s}^d, \mathbf{w})$, is bounded as

$$\begin{aligned} Pr[\hat{\mathcal{K}} = \emptyset] &= Pr[\forall \mathcal{K}' : M2PMI(\mathcal{K}') \leq 0] \\ &\leq Pr[M2PMI(\mathcal{K}) \leq 0] \\ &= Pr[\overset{\circ}{I}(\mathbf{u}_{\mathcal{K}}; \mathbf{y} | \mathbf{s}^d \mathbf{w}) \leq K(\rho(p_{\mathbf{s} | \mathbf{s}^d \mathbf{w}}) + R + \Delta)] \\ &\stackrel{(a)}{=} \exp_2 \left\{ -N \check{E}_{psp, \mathcal{K}, N}(R + \Delta, L_w, L_u, p_{\mathbf{w}}^*, p_{\mathbf{s} | \mathbf{w}}, p_{\mathbf{xu} | \mathbf{sw}}^*, \mathcal{W}_K) \right\}. \end{aligned} \quad (\text{B.21})$$

where (a) follows from (C.5) with $\nu = R + \Delta$. To bound $Pr[\hat{\mathcal{K}} = \mathcal{I}]$, we use property (5.5) with $\hat{\mathcal{K}} = \mathcal{I}$ and $\mathcal{A} = \mathcal{K}$, which yields

$$\overset{\circ}{I}(\mathbf{u}_{\mathcal{K}}; \mathbf{y}_{\mathcal{I}} | \mathbf{s}^d \mathbf{w}) \leq K(\rho(p_{\mathbf{s} | \mathbf{s}^d \mathbf{w}}) + R + \Delta).$$

Since

$$\overset{\circ}{I}(\mathbf{u}_{\mathcal{K}}; \mathbf{y}_{\mathcal{I}} | \mathbf{s}^d \mathbf{w}) = \overset{\circ}{I}(\mathbf{u}_{\mathcal{K}}; \mathbf{y} | \mathbf{s}^d \mathbf{w}) + I(\mathbf{u}_{\mathcal{K}}; \mathbf{u}_{\mathcal{I}} | \mathbf{y} \mathbf{s}^d \mathbf{w}) \geq \overset{\circ}{I}(\mathbf{u}_{\mathcal{K}}; \mathbf{y} | \mathbf{s}^d \mathbf{w})$$

combining the two inequalities above yields

$$\overset{\circ}{I}(\mathbf{u}_{\mathcal{K}}; \mathbf{y} | \mathbf{s}^d \mathbf{w}) \leq K(\rho(p_{\mathbf{s} | \mathbf{s}^d \mathbf{w}}) + R + \Delta).$$

The probability of this event is again given by (B.21); we conclude that

$$p_{miss-all}(p_{\mathbf{w}}^* p_{\mathbf{s} | \mathbf{w}}, p_{\mathbf{xu} | \mathbf{sw}}^*, \mathcal{W}_K) \doteq \exp_2 \left\{ -N \check{E}_{psp, \mathcal{K}, N}(R + \Delta, L_w, L_u, p_{\mathbf{w}}^*, p_{\mathbf{s} | \mathbf{w}}, p_{\mathbf{xu} | \mathbf{sw}}^*, \mathcal{W}_K) \right\}.$$

Averaging over \mathbf{S} and proceeding as in Part (ii) above, we obtain

$$\begin{aligned} p_{miss-all}(\mathcal{W}_K) &\leq \sum_{p_{\mathbf{s} | \mathbf{w}}} Pr[T_{\mathbf{s} | \mathbf{w}}] p_{miss-all}(p_{\mathbf{w}}^* p_{\mathbf{s} | \mathbf{w}}, p_{\mathbf{xu} | \mathbf{sw}}^*, \mathcal{W}_K) \\ &\doteq \exp_2 \left\{ -N \bar{E}_{psp}(R + \Delta, L_w, L_u, D_1, \mathcal{W}_K) \right\} \end{aligned}$$

which establishes (5.15).

(iv). Optimal Collusion Channels are Fair. The proof parallels that of [10, Theorem 4.1(iv)] and is omitted.

(v). Detect-All Exponent for Fair Collusion Channels. The proof parallels that of [10, Theorem 4.1(v)] and is omitted.

(vi). Achievable Rates. Consider any $\mathcal{W} = \{1, \dots, L_w\}$ and p_W that is positive over its support set (if it is not, reduce the value of L_w accordingly.) For any $\mathcal{A} \subseteq \mathcal{K}$, the divergence to be minimized in the expression (5.7) for $\tilde{E}_{psp, \mathcal{A}}(R, L_w, L_u, p_W, \tilde{p}_{S|W}, p_{XU|SW}, \mathcal{W}_K)$ is zero if and only if

$$\tilde{p}_{Y(XU)_{\mathcal{K}}|SW} = \tilde{p}_{Y|X_{\mathcal{K}}} p_{XU|SW}^{\mathcal{K}} \quad \text{and} \quad \tilde{p}_{S|W} = p_S.$$

These p.m.f.'s are feasible for (5.6) if and only if the inequality below holds:

$$\frac{1}{|\mathcal{A}|} I(U_{\mathcal{A}}; YU_{\mathcal{K} \setminus \mathcal{A}} | S^d, W) > I(U; S | S^d, W) + R.$$

They are infeasible, and thus positive error exponents are guaranteed, if

$$R < \min_{\mathcal{A} \subseteq \mathcal{K}} \frac{1}{|\mathcal{A}|} I(U_{\mathcal{A}}; YU_{\mathcal{K} \setminus \mathcal{A}} | S^d, W) - I(U; S | S^d, W).$$

From Part (iv) above, we may restrict our attention to $\mathcal{W}_K = \mathcal{W}_K^{fair}$ under the detect-one criterion. Since the p.m.f. of $(S, W, (XU)_{\mathcal{K}}, Y)$ is permutation-invariant, by application of [10, Eqn. (3.3)] with $(U_{\mathcal{K}}, S^d)$ in place of $(X_{\mathcal{K}}, S)$, we have

$$\min_{\mathcal{A} \subseteq \mathcal{K}} \frac{1}{|\mathcal{A}|} I(U_{\mathcal{A}}; YU_{\mathcal{K} \setminus \mathcal{A}} | S^d, W) = \frac{1}{K} I(U_{\mathcal{K}}; Y | S^d, W). \quad (\text{B.22})$$

Hence the supremum of all R for error exponents are positive is given by $\underline{C}^{one}(D_1, \mathcal{W}_K)$ in (5.16) and is obtained by letting $\epsilon \rightarrow 0$, $\Delta \rightarrow 0$ and $L_w, L_u \rightarrow \infty$.

For any \mathcal{W}_K , under the detect-all criterion, the supremum of all R for which error exponents are positive is given by $\underline{C}^{all}(D_1, \mathcal{W}_K)$ in (5.17) and is obtained by letting $\epsilon \rightarrow 0$, $\Delta \rightarrow 0$ and $L_w, L_u \rightarrow \infty$. Since the optimal conditional p.m.f. is not necessarily permutation-invariant, (B.22) does not hold in general. However, if $\mathcal{W}_K = \mathcal{W}_K^{fair}$, (B.22) holds, and the same achievable rate is obtained for the detect-one and detect-all problems. \square

APPENDIX III

Lemma 3.1: 1) Fix (s^d, \mathbf{w}) and $\mathbf{z} \in \mathcal{Z}^N$, and draw $\mathbf{u}_{\mathcal{K}} = \{\mathbf{u}_m, m \in \mathcal{K}\}$ i.i.d. uniformly over a common type class $T_{\mathbf{u}|\mathbf{s}^d \mathbf{w}}$, independently of \mathbf{z} . We have the asymptotic equality

$$Pr[T_{\mathbf{u}_{\mathcal{K}}|\mathbf{z} \mathbf{s}^d \mathbf{w}}] = \frac{|T_{\mathbf{u}_{\mathcal{K}}|\mathbf{z} \mathbf{s}^d \mathbf{w}}|}{|T_{\mathbf{u}|\mathbf{s}^d \mathbf{w}}|^K} \doteq 2^{-N[KH(\mathbf{u}|\mathbf{s}^d \mathbf{w}) - H(\mathbf{u}_{\mathcal{K}}|\mathbf{z} \mathbf{s}^d \mathbf{w})]} = 2^{-N\overset{\circ}{I}(\mathbf{u}_{\mathcal{K}}; \mathbf{z}|\mathbf{s}^d \mathbf{w})} \quad (\text{C.1})$$

$$Pr[\overset{\circ}{I}(\mathbf{u}_{\mathcal{K}}; \mathbf{z}|\mathbf{s}^d \mathbf{w}) \geq \nu] \doteq 2^{-N\nu}. \quad (\text{C.2})$$

2) Given \mathbf{w} , draw \mathbf{s} i.i.d. p_S . We have [21]

$$Pr[T_{\mathbf{s}|\mathbf{w}}] \doteq 2^{-ND(p_{\mathbf{s}|\mathbf{w}} \| p_S | p_{\mathbf{w}})}. \quad (\text{C.3})$$

3) Given (\mathbf{s}, \mathbf{w}) , draw $(\mathbf{x}_k, \mathbf{u}_k)$, $k \in \mathcal{K}$, i.i.d. uniformly from a conditional type class $T_{\mathbf{xu}|\mathbf{sw}}$, and then draw \mathbf{Y} uniformly from a single conditional type class $T_{\mathbf{y}|\mathbf{x}_\mathcal{K}}$. We have

$$\begin{aligned} Pr[T_{\mathbf{y}(\mathbf{xu})_\mathcal{K}|\mathbf{sw}}] &= \frac{|T_{\mathbf{y}(\mathbf{xu})_\mathcal{K}|\mathbf{sw}}|}{|T_{\mathbf{y}|\mathbf{x}_\mathcal{K}}|} \frac{|T_{(\mathbf{xu})_\mathcal{K}|\mathbf{sw}}|}{|T_{\mathbf{xu}|\mathbf{sw}}|^K} \\ &\doteq \exp_2 \left\{ -ND(p_{\mathbf{y}|\mathbf{x}_\mathcal{K}|\mathbf{sw}} \| p_{\mathbf{y}|\mathbf{x}_\mathcal{K}} p_{\mathbf{xu}|\mathbf{sw}}^K | p_{\mathbf{sw}}) \right\}. \end{aligned} \quad (\text{C.4})$$

For any feasible, strongly exchangeable collusion channel, for any $\mathcal{A} \subseteq \mathcal{K}$ and $\nu > 0$, we have

$$\begin{aligned} Pr[\overset{\circ}{I}(\mathbf{u}_\mathcal{A}; \mathbf{y} \mathbf{u}_{\mathcal{K} \setminus \mathcal{A}} | \mathbf{s}^d \mathbf{w}) \leq |\mathcal{A}|(\nu + \rho(p_{\mathbf{s}|\mathbf{s}^d \mathbf{w}}))] \\ \doteq \exp_2 \left\{ -N \check{E}_{psp, \mathcal{A}, N}(\nu, L, p_{\mathbf{w}}^*, p_{\mathbf{s}|\mathbf{w}}, p_{\mathbf{xu}|\mathbf{sw}}^*, \mathcal{W}_K) \right\}. \end{aligned} \quad (\text{C.5})$$

Proof: The derivation of (C.4), (C.3), and (C.5) parallels that of (D.12), (D.15) and (D.16) in [10].

APPENDIX IV

PROOF OF THEOREM 6.1

Let K be size of the coalition and (f_N, g_N) a sequence of length- N , rate- R randomized codes. We show that for any sequence of such codes, reliable decoding of all K fingerprints is possible only if $R \leq \bar{C}^{all}(D_1, \mathcal{W}_K)$. Recall that the encoder generates marked copies $\mathbf{x}_m = f_N(\mathbf{s}, v, m)$ for $1 \leq m \leq 2^{NR}$ and that the decoder outputs an estimated coalition $g_N(\mathbf{y}, \mathbf{s}^d, v) \in \{1, \dots, 2^{NR}\}^*$. We use the notation $M^K \triangleq \{M_1, \dots, M_K\}$ and $\mathbf{X}^K \triangleq \{\mathbf{X}_1, \dots, \mathbf{X}_K\}$.

To prove that $\bar{C}^{all}(D_1, \mathcal{W}_K)$ is an upper bound on capacity, it suffices to identify a family of collusion channels for which reliable decoding is impossible at rates above $\bar{C}^{all}(D_1, \mathcal{W}_K)$. As shown in [10], it is sufficient to derive such a bound for the compound family \mathcal{W}_K of *memoryless channels*.

Our derivation is an extension of the single-user compound Gel'fand-Pinsker problem [11] to the multiple-access case. A lower bound on error probability is obtained when an oracle informs the decoder that the coalition size is *at most* K .

There are $\binom{2^{NR}}{K} \leq 2^{KNR}$ possible coalitions of size $\leq K$. We represent such a coalition as $M^K \triangleq \{M_1, \dots, M_K\}$, where M_k , $1 \leq k \leq K$, are drawn i.i.d. uniformly from $\{1, \dots, 2^{NR}\}$.

Given a memoryless channel $p_{Y|X^K} \in \mathcal{W}_K$, the joint p.m.f. of $(M^K, V, \mathbf{S}, \mathbf{X}^K, \mathbf{Y})$ is given by

$$p_{M^K V \mathbf{S} \mathbf{X}^K \mathbf{Y}} = p_S^N p_V \prod_{1 \leq k \leq K} (p_{M_k} \mathbb{1}_{\{\mathbf{x}_k = f_N(\mathbf{s}, V, M_k)\}}) p_{Y|X^K}^N. \quad (\text{D.1})$$

Our derivations make repeated use of the identity

$$I(U_\mathcal{A}; Y | Z, U_{\mathcal{K} \setminus \mathcal{A}}) - I(U_\mathcal{A}; S | Z, U_{\mathcal{K} \setminus \mathcal{A}}) = I(U_\mathcal{A}; Y, Z | U_{\mathcal{K} \setminus \mathcal{A}}) - I(U_\mathcal{A}; S, Z | U_{\mathcal{K} \setminus \mathcal{A}})$$

which follows from the chain rule for conditional mutual information and holds for any (U_K, S, Y, Z) .

The total error probability (including false positives and false negatives) for the detect-all decoder is

$$P_e(p_{Y|X^K}) = Pr[\hat{\mathcal{K}} \neq \mathcal{K}] \quad (\text{D.2})$$

when collusion channel $p_{Y|X^K} \in \mathcal{W}_K$ is in effect.

Step 1. Following the derivation of [10, Eqn. (B.20)] with $(\mathbf{Y}, \mathbf{S}^d, V)$ in place of $(\mathbf{Y}, \mathbf{S}, V)$ at the receiver, for the error probability $P_e(p_{Y|X^K})$ to vanish for each $p_{Y|X^K} \in \mathcal{W}_K$, we need

$$R \leq \liminf_{N \rightarrow \infty} \min_{p_{Y|X^K} \in \mathcal{W}_K} \min_{\mathcal{A} \subseteq \mathcal{K}} \frac{1}{N|\mathcal{A}|} I(M_{\mathcal{A}}; \mathbf{Y} | \mathbf{S}^d, V). \quad (\text{D.3})$$

Step 2. Define the i.i.d. random variables

$$W_i = \{V, S_j, j \neq i\} \in \mathcal{V}_N \times \mathcal{S}^{N-1}, \quad 1 \leq i \leq N. \quad (\text{D.4})$$

Also define the random variables

$$\begin{aligned} V_{ki} &= (M_k, V, S_{i+1}^N), \\ U_{ki} &= (V_{ki}, (Y S^d)^{i-1}) = (M_k, V, S_{i+1}^N, (Y S^d)^{i-1}), \quad 1 \leq k \leq K, 1 \leq i \leq N \end{aligned} \quad (\text{D.5})$$

where $S_{i+1}^N \triangleq (S_{i+1}, \dots, S_N)$ and $(Y S^d)^{i-1} \triangleq (Y_1, S_1^d, \dots, Y_{i-1}, S_{i-1}^d)$. Hence

$$V_{i-1}^K = (V_i^K, S_i), \quad V_1^K = U_1^K, \quad V_N^K = (M^K, V). \quad (\text{D.6})$$

The following properties hold for each $1 \leq i \leq N$:

- By (D.1) and (D.5), $(S_i, W_i, U_i^K) = (M^K, V, \mathbf{S}, Y^{i-1}) \rightarrow X_i^K \rightarrow Y_i$ forms a Markov chain.
- The random variables X_{ki} , $1 \leq k \leq K$, are conditionally i.i.d. given $(\mathbf{S}, V) = (S_i, W_i)$.
- Due to the term Y^{i-1} in (D.5), the random variables U_{ki} , $1 \leq k \leq K$, are conditionally *dependent* given $(\mathbf{S}, V) = (S_i, W_i)$.

The joint p.m.f. of $(S_i, W_i, X_i^K, U_i^K, Y_i)$ may thus be written as

$$p_{S_i} p_{W_i} \left(\prod_{1 \leq k \leq K} p_{X_{ki} | S_i W_i} \right) p_{U_i^K | X_i^K S_i W_i} p_{Y_i | X_K}, \quad 1 \leq i \leq N. \quad (\text{D.7})$$

Step 3. Consider a time-sharing random variable T that is uniformly distributed over $\{1, \dots, N\}$ and independent of the other random variables, and define the tuple of random variables (S, S^d, W, U^K, X^K, Y) as $(S_T, S_T^d, W_T, U_T^K, X_T^K, Y_T)$. Also let $W = (W_T, T)$ and $U_k = (U_{k,T}, T)$, $1 \leq k \leq K$, which are defined over alphabets of respective cardinalities

$$L_w(N) = N |\mathcal{V}_N| |\mathcal{S}|^{N-1}$$

and

$$L_u(N) = N |\mathcal{V}_N| 2^{N[R + \log \max(|\mathcal{S}|, |\mathcal{Y}| |S^d|)]}.$$

Since $(S_i, W_i, U_i^K) \rightarrow X_i^K \rightarrow Y_i$ forms a Markov chain, so does $(S, W, U^K) \rightarrow X^K \rightarrow Y$. From (D.7), the joint p.m.f. of (S, W, U^K, X^K, Y) takes the form

$$p_S p_W \left(\prod_{1 \leq k \leq K} p_{X_k|SW} \right) p_{U^K|X^K SW} p_{Y|X^K}. \quad (\text{D.8})$$

In (6.1) we have defined the set

$$\begin{aligned} \mathcal{P}_{X^K U^K W|S}^{\text{outer}}(p_S, L_w, L_u, D_1) = & \left\{ p_{X^K U^K W|S} = p_W \left(\prod_{k=1}^K p_{X_k|SW} \right) p_{U^K|X^K SW} \right. \\ & \left. : p_{X_1|SW} = \dots = p_{X_K|SW}, \text{ and } \mathbb{E}d(S, X_1) \leq D_1 \right\} \end{aligned} \quad (\text{D.9})$$

where $|\mathcal{W}| = L_w$ and $|\mathcal{U}| = L_u$. Observe that $p_{X^K U^K W|S}$ defined in (D.8) belongs to $\mathcal{P}_{X^K U^K W|S}(p_S, L_w, L_u, D_1)$.

Define the collection of K indices $\mathcal{K} = \{1, 2, \dots, K\}$ and the following functionals indexed by $\mathcal{A} \subseteq \mathcal{K}$:

$$J_{L_w, L_u, \mathcal{A}}(p_S, p_{X^K U^K W|S}, p_{Y|X^K}) = \frac{1}{|\mathcal{A}|} [I(U_{\mathcal{A}}; Y S^d | U_{\mathcal{K} \setminus \mathcal{A}}) - I(U_{\mathcal{A}}; S | U_{\mathcal{K} \setminus \mathcal{A}})]. \quad (\text{D.10})$$

Step 4. We have

$$\begin{aligned} I(M_{\mathcal{K}}; \mathbf{Y} | \mathbf{S}^d, V) & \stackrel{(a)}{=} I(M_{\mathcal{K}}; \mathbf{Y} | \mathbf{S}^d, V) - I(M_{\mathcal{K}}, V; \mathbf{S} | \mathbf{S}^d) \\ & = I(M_{\mathcal{K}}, V; \mathbf{Y} | \mathbf{S}^d) - I(V; \mathbf{Y} | \mathbf{S}^d) - I(M_{\mathcal{K}}, V; \mathbf{S} | \mathbf{S}^d) \\ & \leq I(M_{\mathcal{K}}, V; \mathbf{Y} | \mathbf{S}^d) - I(M_{\mathcal{K}}, V; \mathbf{S} | \mathbf{S}^d) \\ & \stackrel{(b)}{=} I(M_{\mathcal{K}}, V; \mathbf{Y} \mathbf{S}^d) - I(M_{\mathcal{K}}, V; \mathbf{S}) \\ & \stackrel{(c)}{\leq} \sum_{i=1}^N [I(U_{\mathcal{K}, i}; Y_i S_i^d) - I(U_{\mathcal{K}, i}; S_i)] \\ & = I(U_{\mathcal{K}, T}; Y S^d | T) - I(U_{\mathcal{K}, T}; S | T) \\ & = I(U_{\mathcal{K}, T}, T; Y S^d) - I(T; Y S^d) - I(U_{\mathcal{K}, T}, T; S) + I(T; S) \\ & \stackrel{(d)}{\leq} I(U_{\mathcal{K}, T}, T; Y S^d) - I(U_{\mathcal{K}, T}, T; S) \\ & \stackrel{(e)}{=} I(U_{\mathcal{K}}; Y S^d) - I(U_{\mathcal{K}}; S) \\ & = K J_{L_w(N), L_u(N), \mathcal{K}}(p_S, p_{X^K U^K W|S}, p_{Y|X^K}), \end{aligned} \quad (\text{D.11})$$

where (a) holds because $M_{\mathcal{K}}, V, \mathbf{S}$ are mutually independent, and (b) follows from the chain rule for mutual information, (c) from [20, Lemma 4], using V_i^K and U_i^K in place of V_i and U_i , respectively, (d) holds because $I(T; S) = 0$, and (e) by definition of $U_{\mathcal{K}}$.

For all $\mathcal{A} \subset \mathcal{K}$, we have

$$\begin{aligned}
I(M_{\mathcal{A}}; \mathbf{Y} | \mathbf{S}^d, V) &= I(M_{\mathcal{A}}, V; \mathbf{Y} | \mathbf{S}^d, V) \\
&\stackrel{(a)}{=} I(M_{\mathcal{A}}, V; \mathbf{Y} | \mathbf{S}^d, V) - I(M_{\mathcal{A}}, V; \mathbf{S} | \mathbf{S}^d, M_{\mathcal{K} \setminus \mathcal{A}}, V) \\
&\stackrel{(b)}{=} I(M_{\mathcal{A}}, V; \mathbf{Y} | \mathbf{S}^d, M_{\mathcal{K} \setminus \mathcal{A}}, V) - I(M_{\mathcal{A}}, V; \mathbf{S} | \mathbf{S}^d, M_{\mathcal{K} \setminus \mathcal{A}}, V) \\
&= I(M_{\mathcal{A}}, V; \mathbf{Y} \mathbf{S}^d | M_{\mathcal{K} \setminus \mathcal{A}}, V) - I(M_{\mathcal{A}}, V; \mathbf{S} | \mathbf{S}^d, M_{\mathcal{K} \setminus \mathcal{A}}, V) \\
&\stackrel{(c)}{=} \sum_{i=1}^N [I(U_{\mathcal{A}, i}; Y_i S_i^d | U_{\mathcal{K} \setminus \mathcal{A}, i}) - I(U_{\mathcal{A}, i}; S_i | U_{\mathcal{K} \setminus \mathcal{A}, i})] \\
&= N [I(U_{\mathcal{A}, T}; Y S^d | U_{\mathcal{K} \setminus \mathcal{A}, T}, T) - I(U_{\mathcal{A}, T}; S | U_{\mathcal{K} \setminus \mathcal{A}, T}, T)] \\
&= N [I(U_{\mathcal{A}, T}, T; Y S^d | U_{\mathcal{K} \setminus \mathcal{A}, T}, T) - I(U_{\mathcal{A}, T}, T; S | U_{\mathcal{K} \setminus \mathcal{A}, T}, T)] \\
&\stackrel{(d)}{=} N [I(U_{\mathcal{A}}; Y S^d | U_{\mathcal{K} \setminus \mathcal{A}}) - I(U_{\mathcal{A}}; S | U_{\mathcal{K} \setminus \mathcal{A}})] \\
&= N |\mathcal{A}| J_{L_w(N), L_u(N), \mathcal{A}}(p_S, p_{X^K U^K W} | S, p_{Y | X^K}). \tag{D.13}
\end{aligned}$$

where (a) and (b) hold because M_K , \mathbf{S} , and V are mutually independent, the equality (c) is proved at the end of this section, and (d) follows from the definition of U_K .

Combining (D.3), (D.11), and (D.13), we obtain

$$\begin{aligned}
R &\leq \liminf_{N \rightarrow \infty} \min_{p_{Y | X^K} \in \mathcal{W}_K} \min_{\mathcal{A} \subseteq \mathcal{K}} J_{L_w(N), L_u(N), \mathcal{A}}(p_S, p_{X^K U^K W} | S, p_{Y | X^K}) \\
&\stackrel{(a)}{\leq} \sup_{L_w, L_u} \min_{p_{Y | X^K} \in \mathcal{W}_K} \min_{\mathcal{A} \subseteq \mathcal{K}} J_{L_w, L_u, \mathcal{A}}(p_S, p_{X^K U^K W} | S, p_{Y | X^K}) \\
&\leq \sup_{L_w, L_u} \max_{p_{X^K U^K W} | S \in \mathcal{P}_{X^K U^K W} | S(p_S, L_w, L_u, D_1)} \min_{p_{Y | X^K} \in \mathcal{W}_K} \min_{\mathcal{A} \subseteq \mathcal{K}} J_{L_w, L_u, \mathcal{A}}(p_S, p_{X^K U^K W} | S, p_{Y | X^K}) \\
&\stackrel{(b)}{=} \sup_{L_w, L_u} \overline{C}_{L_w, L_u}^{all}(D_1, \mathcal{W}_K) \\
&= \lim_{L_w, L_u \rightarrow \infty} \overline{C}_{L_w, L_u}^{all}(D_1, \mathcal{W}_K) \\
&\stackrel{(c)}{=} \overline{C}^{all}(D_1, \mathcal{W}_K), \tag{D.14}
\end{aligned}$$

where (a) holds because the functionals $J_{L_w, L_u, \mathcal{A}}(\cdot)$ are nondecreasing in L_w, L_u , (b) uses the definition of $\overline{C}_{L_w, L_u}^{all}$ in (6.2), and (c) the fact that the sequence $\{\overline{C}_{L_w, L_u}^{all}\}$ is nondecreasing.

Proof of (D.12). Recall the definitions of $V_{\mathcal{K},i} = (M_{\mathcal{K}}, V, S_{i+1}^N)$ and $U_{\mathcal{K},i} = (V_{\mathcal{K},i}, (YS^d)^{i-1})$ in (D.5) and the recursion (D.6) for $V_{\mathcal{K},i}$. We prove the following inequality:

$$\begin{aligned} & I(U_{\mathcal{A},i}; Y_i S_i^d | U_{\mathcal{K} \setminus \mathcal{A},i}) - I(U_{\mathcal{A},i}; S_i | U_{\mathcal{K} \setminus \mathcal{A},i}) \\ &= [I(V_{\mathcal{A},i}; (YS^d)^i | V_{\mathcal{K} \setminus \mathcal{A},i}) - I(V_{\mathcal{A},i}; S_i^i | V_{\mathcal{K} \setminus \mathcal{A},i})] \\ &\quad - [I(V_{\mathcal{A},i-1}; (YS^d)^{i-1} | V_{\mathcal{K} \setminus \mathcal{A},i-1}) - I(V_{\mathcal{A},i-1}; S_i^{i-1} | V_{\mathcal{K} \setminus \mathcal{A},i-1})]. \end{aligned} \quad (\text{D.15})$$

Then summing both sides of this equality from $i = 2$ to N , cancelling terms, and using the properties $V_{k,1} = U_{k,1}$ and $V_{k,N} = (M_k, V)$ yields (D.12).

The first of the six terms in (D.15) may be expanded as follows:

$$\begin{aligned} I(U_{\mathcal{A},i}; Y_i S_i^d | U_{\mathcal{K} \setminus \mathcal{A},i}) &= I(V_{\mathcal{A},i}, (YS^d)^{i-1}; Y_i S_i^d | V_{\mathcal{K} \setminus \mathcal{A},i}, (YS^d)^{i-1}) \\ &= I(V_{\mathcal{A},i}; Y_i S_i^d | V_{\mathcal{K} \setminus \mathcal{A},i}, (YS^d)^{i-1}) \\ &= I(V_{\mathcal{A},i}, (YS^d)^{i-1}; Y_i S_i^d | V_{\mathcal{K} \setminus \mathcal{A},i}) - I((YS^d)^{i-1}; Y_i S_i^d | V_{\mathcal{K} \setminus \mathcal{A},i}) \\ &= I(U_{\mathcal{A},i}; Y_i S_i^d | V_{\mathcal{K} \setminus \mathcal{A},i}) - I((YS^d)^{i-1}; Y_i S_i^d | V_{\mathcal{K} \setminus \mathcal{A},i}). \end{aligned} \quad (\text{D.16})$$

Similarly for the second term, replacing (YS^d) with S in the above derivation, we obtain

$$I(U_{\mathcal{A},i}; S_i | U_{\mathcal{K} \setminus \mathcal{A},i}) = I(U_{\mathcal{A},i}; S_i | V_{\mathcal{K} \setminus \mathcal{A},i}) - I((YS^d)^{i-1}; S_i | V_{\mathcal{K} \setminus \mathcal{A},i}). \quad (\text{D.17})$$

The six terms in (D.15) can be expanded using the chain rule for mutual information, in the same way as in [20, Lemma 4.2]:

$$I(V_{\mathcal{A},i}; (YS^d)^i | V_{\mathcal{K} \setminus \mathcal{A},i}) = I(V_{\mathcal{A},i}; (YS^d)^{i-1} | V_{\mathcal{K} \setminus \mathcal{A},i}) + I(V_{\mathcal{A},i}; (YS^d)_i | V_{\mathcal{K} \setminus \mathcal{A},i}) \quad (\text{D.18})$$

$$I(V_{\mathcal{A},i}; S_i^i | V_{\mathcal{K} \setminus \mathcal{A},i}) = I(V_{\mathcal{A},i}; S_i^{i-1} | V_{\mathcal{K} \setminus \mathcal{A},i}) + I(V_{\mathcal{A},i}; S_i | V_{\mathcal{K} \setminus \mathcal{A},i}) \quad (\text{D.19})$$

$$I(V_{\mathcal{A},i-1}; S_i^{i-1} | V_{\mathcal{K} \setminus \mathcal{A},i-1}) = I(V_{\mathcal{A},i}; S_i^{i-1} | S_i, V_{\mathcal{K} \setminus \mathcal{A},i-1}) \quad (\text{D.20})$$

$$I(V_{\mathcal{A},i-1}; (YS^d)^{i-1} | V_{\mathcal{K} \setminus \mathcal{A},i-1}) = I(V_{\mathcal{A},i}; (YS^d)^{i-1} | S_i, V_{\mathcal{K} \setminus \mathcal{A},i-1}) \quad (\text{D.21})$$

$$I(U_{\mathcal{A},i}; S_i | V_{\mathcal{K} \setminus \mathcal{A},i}) = I((YS^d)^{i-1}; S_i | V_{\mathcal{K} \setminus \mathcal{A},i}) + I(V_{\mathcal{A},i}; S_i | (YS^d)^{i-1}, V_{\mathcal{K} \setminus \mathcal{A},i}) \quad (\text{D.22})$$

$$\begin{aligned} I(U_{\mathcal{A},i}; (YS^d)_i | V_{\mathcal{K} \setminus \mathcal{A},i}) &= I((YS^d)^{i-1}; (YS^d)_i | V_{\mathcal{K} \setminus \mathcal{A},i}) + I(V_{\mathcal{A},i}; (YS^d)_i | (YS^d)^{i-1}, V_{\mathcal{K} \setminus \mathcal{A},i}). \end{aligned} \quad (\text{D.23})$$

Moreover, expanding the conditional mutual information $I(V_{\mathcal{A},i}; S_i, (YS^d)^{i-1} | V_{\mathcal{K} \setminus \mathcal{A},i})$ in two different ways, we obtain

$$\begin{aligned} & I(V_{\mathcal{A},i}; (YS^d)^{i-1} | V_{\mathcal{K} \setminus \mathcal{A},i}) + I(V_{\mathcal{A},i}; S_i | (YS^d)^{i-1}, V_{\mathcal{K} \setminus \mathcal{A},i}) \\ &= I(V_{\mathcal{A},i}; S_i^{i-1} | V_{\mathcal{K} \setminus \mathcal{A},i}) + I(V_{\mathcal{A},i}; (YS^d)^{i-1} | S_i, V_{\mathcal{K} \setminus \mathcal{A},i}). \end{aligned} \quad (\text{D.24})$$

Subtracting the sum of (D.17), (D.18), (D.20), (D.22), (D.24) from the sum of (D.16), (D.19), (D.21), (D.23), and cancelling terms, we obtain (D.15), from which the claim follows. \square

REFERENCES

- [1] D. Boneh and J. Shaw, “Collusion-Secure Fingerprinting for Digital Data,” in *Advances in Cryptology: Proc. CRYPTO’95*, Springer-Verlag, New York, 1995.
- [2] I. J. Cox, J. Killian, F. T. Leighton and T. Shamoon, “Secure Spread Spectrum Watermarking for Multimedia,” *IEEE Trans. Image Proc.*, Vol. 6, No. 12, pp. 1673–1687, Dec. 1997.
- [3] M. Wu, W. Trappe, Z. J. Wang and K. J. R. Liu, “Collusion-Resistant Fingerprinting for Multimedia,” *IEEE Signal Processing Magazine*, Vol. 21, No. 2, pp. 15–27, March 2004.
- [4] K. J. R. Liu, W. Trappe, Z. J. Wang, M. Wu and H. Zhao, *Multimedia Fingerprinting Forensics for Traitor Tracing*, EURASIP Book Series on Signal Processing, 2006.
- [5] P. Moulin and A. Briassouli, “The Gaussian Fingerprinting Game,” *Proc. Conf. Information Sciences and Systems*, Princeton, NJ, March 2002.
- [6] P. Moulin and J. A. O’Sullivan, “Information-theoretic analysis of information hiding,” *IEEE Trans. on Information Theory*, Vol. 49, No. 3, pp. 563–593, March 2003.
- [7] A. Somekh-Baruch and N. Merhav, “On the capacity game of private fingerprinting systems under collusion attacks,” *IEEE Trans. Information Theory*, vol. 51, no. 3, pp. 884–899, Mar. 2005.
- [8] A. Somekh-Baruch and N. Merhav, “Achievable error exponents for the private fingerprinting game,” *IEEE Trans. Information Theory*, Vol. 53, No. 5, pp. 1827–1838, May 2007.
- [9] N. P. Anthapadmanabhan, A. Barg and I. Dumer, “On the Fingerprinting Capacity Under the Marking Assumption,” submitted to *IEEE Trans. Information Theory*, arXiv:cs/0612073v2, July 2007.
- [10] P. Moulin, “Universal Fingerprinting: Capacity and Random-Coding Exponents,” submitted to *IEEE Trans. Information Theory*. Available from arXiv:0801.3837v1 [cs.IT] 24 Jan 2008.
- [11] P. Moulin and Y. Wang, “Capacity and Random-Coding Exponents for Channel Coding with Side Information,” *IEEE Trans. on Information Theory*, Vol. 53, No. 4, pp. 1326–1347, Apr. 2007.
- [12] Y.-S. Liu and B. L. Hughes, “A new universal random coding bound for the multiple-access channel,” *IEEE Trans. Information Theory*, vol. 42, no. 2, pp. 376–386, Mar. 1996.
- [13] A. Somekh-Baruch and N. Merhav, “On the Random Coding Error Exponents of the Single-User and the Multiple-Access Gel’fand-Pinsker Channels,” *Proc. IEEE Int. Symp. Info. Theory*, p. 448, Chicago, IL, June-July 2004.
- [14] Y. Wang and P. Moulin, “Capacity and Random-Coding Error Exponent for Public Fingerprinting Game,” *Proc. Int. Symp. on Information Theory*, Seattle, WA, July 2006.
- [15] Y. Wang, *Detection- and Information-Theoretic Analysis of Steganography and Fingerprinting*, Ph. D. Thesis, ECE Department, University of Illinois at Urbana-Champaign, Dec. 2006.
- [16] G. Tardos, “Optimal Probabilistic Fingerprinting Codes,” *STOC*, 2003.
- [17] R. Ahlswede, “Multiway Communication Channels,” *Proc. ISIT*, pp. 23–52, Tsahkadsor, Armenia, 1971.
- [18] H. Liao, “Multiple Access Channels,” *Ph. D. dissertation*, EE Department, U. of Hawaii, 1972.
- [19] A. Lapidoth and P. Narayan, “Reliable Communication Under Channel Uncertainty,” *IEEE Trans. Information Theory*, Vol. 44, No. 6, pp. 2148–2177, Oct. 1998.

- [20] S. I. Gel'fand and M. S. Pinsker, "Coding for Channel with Random Parameters," *Problems of Control and Information Theory*, Vol. 9, No. 1, pp. 19—31, 1980.
- [21] I. Csiszár and J. Körner, *Information Theory: Coding Theory for Discrete Memoryless Systems*, Academic Press, NY, 1981.
- [22] A. Das and P. Narayan, "Capacities of Time-Varying Multiple-Access Channels With Side Information," *IEEE Trans. Information Theory*, Vol. 48, No. 1, pp. 4—25, Jan. 2002.
- [23] S. Sigurjónsson and Y.-H. Kim, "On Multiple User Channels with State Information at the Transmitters," *Proc. ISIT* 2005.
- [24] N. T. Gaarder and J. K. Wolf, "The Capacity Region of a Multiple-Access Discrete Memoryless Channel Can Increase with Feedback," *IEEE Trans. Information Theory*, Vol. 21, No. 1, pp. 100—102, Jan. 1975.